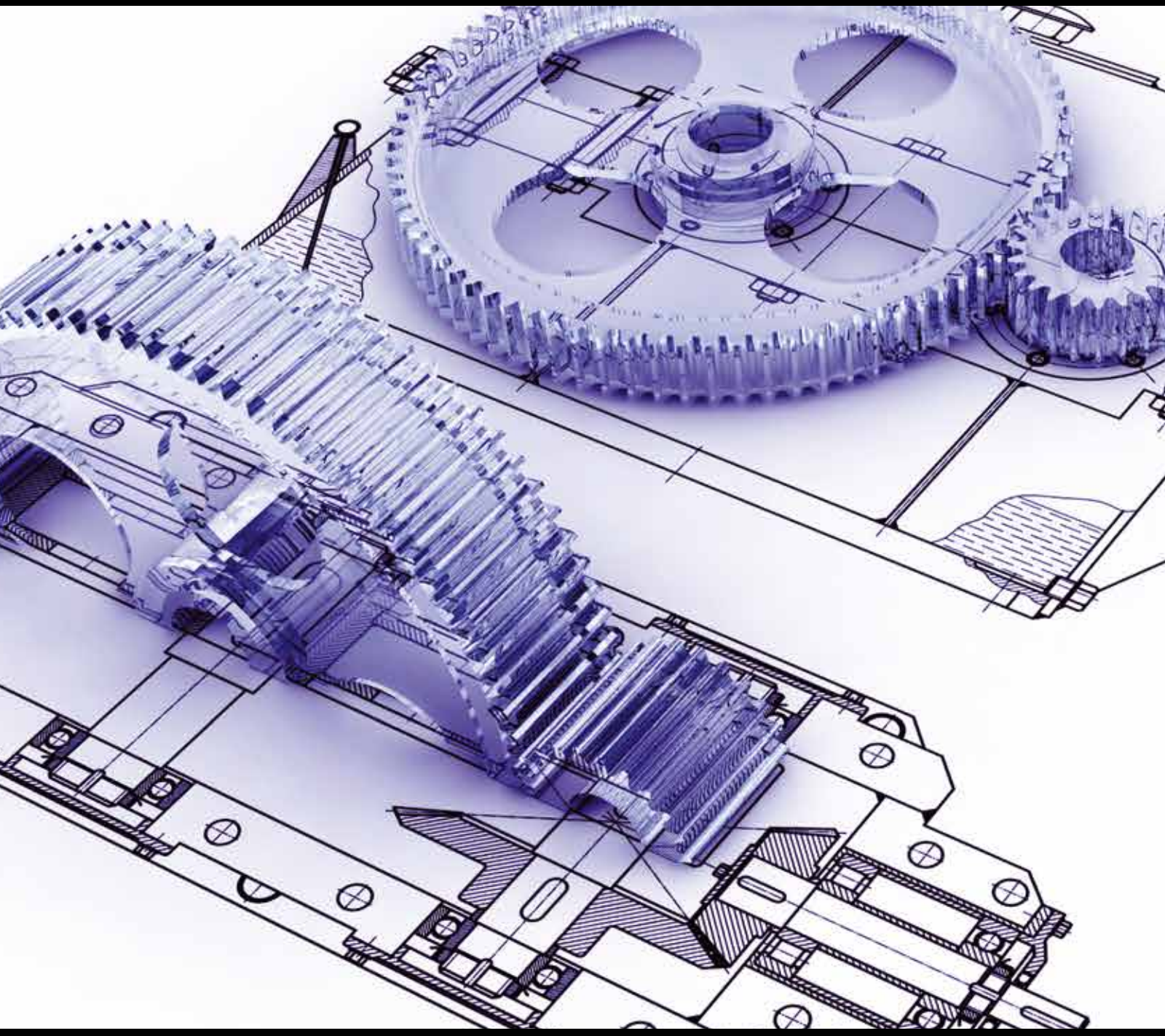
Advances in Mechanical Engineering

Mechatronics and Embedded System

Guest Editors: Yong Tao, Jindong Tan, Zili Shao, and Hongxing Wei



Mechatronics and Embedded System

Advances in Mechanical Engineering

Mechatronics and Embedded System

Guest Editors: Yong Tao, Jindong Tan, Zili Shao, and Hongxing Wei

Editorial Board

Contents

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2014, Article ID 239452, 2 pages
http://dx.doi.org/10.1155/2014/239452

Editorial
Mechatronics and Embedded Systems

Yong Tao,[1] Jindong Tan,[2] Zili Shao,[3] and Hongxing Wei[1]

[1] *School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China*
[2] *The University of Tennessee, Knoxville, TN 37996-2210, USA*
[3] *Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong*

Correspondence should be addressed to Yong Tao; taoyong@me.buaa.edu.cn

Received 24 March 2014; Accepted 24 March 2014; Published 15 May 2014

With the supports for sustainable development of autonomous and intelligent systems, mechatronics and embedded systems are becoming one hot research area. In this special issue, we select 18 papers among a large number of submissions. The selected papers cover several key technology areas in mechatronics and embedded systems and can be divided into the following six categories.

Three papers are related to cutting-edge techniques in robotics. In the paper titled *"Path tracking control of a mobile robot by using dual estimation algorithm,"* a novel dual estimation algorithm is proposed to overcome the slip limitations in the path tracking control of skid-steered mobile robots. In the paper titled *"Self-assembling for swarm modular robots using MIMO fuzzy control,"* a set of fuzzy logic rules and theorems are developed and applied in the controller design of swarm modular robots. In the paper titled *"Development of lifting and propulsion mechanism for biped robot inspired by basilisk lizards,"* the authors developed the lifting and propulsion mechanism of a novel biped robot inspired by the basilisk lizard's water-walking function.

Three papers discuss vision-based obstacle avoidance, navigation, and tracking in robotics. The paper titled *"A hybrid architecture for vision-based obstacle avoidance"* presents a new obstacle avoidance method using a single monocular vision camera as the only sensor. In the paper titled *"Autonomous vehicle navigation using vision and mapless strategies: a survey,"* the authors survey the existing state of knowledge related to vision-based mobile robots, especially including their background and history and current trends in mapless navigation. In the paper titled *"A scale adaptive mean-shift tracking algorithm for robot vision,"* the authors propose a vision tracking algorithm that is well adaptive to scale changing and robust to partial occlusion, and the tracking speed is fast enough for real-time tracking applications in robot vision.

Three papers focus on developing optimization algorithms and frameworks in mechatronics and embedded systems. In the paper titled *"Study on the algorithm for train operation adjustment based on ordinal optimization,"* a railway transportation dispatch optimization mechanism is developed to automatically compile train operation adjustment plan so as to ensure safe, fast, and punctual running of trains. The paper titled *"Extended-maxima transform watershed segmentation algorithm for touching corn kernels"* presents a modified watershed segmentation algorithm that can effectively improve recognition accuracy in corn seed vitality tests. In the paper titled *"A general framework of mechatronic modular architecture,"* the authors propose a framework with the focus on module reuse for modular mechatronic system design.

Three papers target at applications in mechatronics and embedded systems. In the paper titled *"Assessment of the kinematic parameters of distant noncooperative objects,"* the authors propose a method that can assess the kinematic parameters of noncooperative objects. The paper titled *"IPv6-based smart metering network for monitoring building electricity"* presents a smart electricity monitoring system that can reliably and accurately measure and transfer data using ZigBee and IPv6. In the paper titled *"Research on the detection of metal debris with microplane inductance sensor,"* a system is designed to simulate and analyze metal debris.

Three papers present FPGA optimization and implantation techniques in mechatronics and embedded systems. In the paper titled *"Optimal partial reconfiguration for permanent fault recovery on SRAM-based FPGAs in space mission,"* the authors propose a technique to maximize the lifetime of SRAM-based FPGAs in space mission.

The paper titled *"FPGA implementation of real-time Ethernet for motion control"* presents an implementation of real-time Ethernet named CASNET, which modifies the Ethernet medium access control (MAC) to achieve the real-time requirement for motion control. In the paper titled *"Research on optimization for motion control bus based on Ethernet,"* the authors propose an optimization technique to enhance efficiency and reliability of motion control bus of the Ethernet physical layer.

Three papers provide system development in mechatronics and embedded systems. In the paper titled *"Development of a smart residential fire protection system,"* the authors present the development of a smart residential fire detection and extinguishing system. The paper titled *"A force-position control system based on soft tissue under large deformation"* introduces an extension of the mass-spring model that can reflect the mechanical properties of human tissue. In the paper titled *"Look-ahead algorithm with whole s-curve acceleration and deceleration,"* an algorithm is proposed to increase the smoothness of feed rate profile and machining efficiency of continuous short line and circle machining in CNC systems.

Acknowledgments

We would like to take this opportunity to thank the authors for their contributions to this special issue and the thoughtful reviewers.

Yong Tao
Jindong Tan
Zili Shao
Hongxing Wei

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 367127, 11 pages
http://dx.doi.org/10.1155/2013/367127

Research Article

Path Tracking Control of a Mobile Robot by Using Dual Estimation Algorithm

Xuan Vinh Ha,[1] Cheolkeun Ha,[1] and Hyoung Sik Choi[2]

[1] *School of Mechanical Engineering, University of Ulsan, Ulsan 680-749, Republic of Korea*
[2] *Korea Aerospace Research Institute, Daejeon 305-806, Republic of Korea*

Correspondence should be addressed to Cheolkeun Ha; hacheo121c@yahoo.com

Received 12 March 2013; Accepted 7 October 2013

Academic Editor: Hongxing Wei

Skid-steered mobile robots have been developed to explore unknown environments, especially in rough terrain situations, where the wheel slips always occur and vary when the robot is traveling with different trajectory shapes. The wheel slippage limits the traction and braking abilities of the robot. In this paper, we propose a new dual estimation algorithm to overcome the previous limitations. The estimated values such as the robot's positions and wheel slips are obtained based on the Kalman filtering technique. As demonstrated by our experimental results, the advantages of the new method are effective to overcome the slip limitations in the path tracking control problem of a four-track wheel skid-steered mobile robot (4-TW SSMR).

1. Introduction

Recently, several studies have proposed and developed various approaches for estimating the position of a mobile robot. For example, one study used an extended Kalman filter (EKF) with a low-cost GPS unit and inclinometer to develop a localization scheme for Ackerman steering vehicles in indoor autonomous navigation situations by estimating the positions of the vehicles and their sensor biases [1]. In another study, a relative localization method was used to determine the navigational route of a convoy of robotic units in an indoor environment using an EKF based on a low-cost laser range system and built-in odometric sensors [2]. In skid-steered vehicles, discrepancies remain between position measurements and the navigational estimations obtained from odometry, infrared, and ultrasonic measurements and Kalman filtering techniques [3]. However, this study only mentioned the advantages of the different extensions of the KF for nonlinear estimations. In [4], a laser range finder (LRF) was used to find the localization of a mobile robot based on the EKF design. That study focused on a statistically exact derivation of the input noise covariance matrix, according to the known noise variances of the angular velocity measurements of both robot wheels.

Additionally, the exploration of unknown planetary surfaces with the help of mobile robot is increasingly being developed. A localization and path selection of any vehicle like a mobile robot to achieve any planetary exploration missions were described in [5]. In this study, by using odometry and gyro measurements, these sensors information was integrated by EKF. In another research, a robot, named Hyperion, was designed for solar-powered operation in polar environments and developed sun-cognizant navigation method to enable rovers to dodge shadows, seek sun, and drive sun-synchronous routes [6].

In a mobile robot's field of operation, wheel slip limits traction and braking ability, especially for a 4-TW SSMR. Accurate estimation of slip is essential to obtain precise position for a mobile robot operating in unstructured terrain, especially in rough terrain situations. Several approaches to slip estimation have been developed. For example, a sliding mode observer (SMO) conducted to estimate slip parameters based on the kinematic model of a skid-steering vehicle is proposed in [7]. In another study, a novel method combining an optical flow algorithm with a sliding mode observer to estimate slip parameters of a SSMR was introduced [8]. Reference [9] proposed an experimental slip model for exact kinematic modeling, and the parameters of this model were

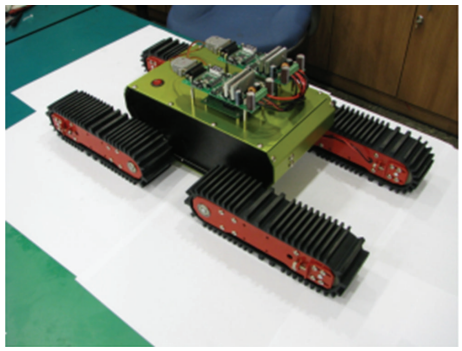


Figure 1: The NT-Hazard Escape-1 mobile robot.

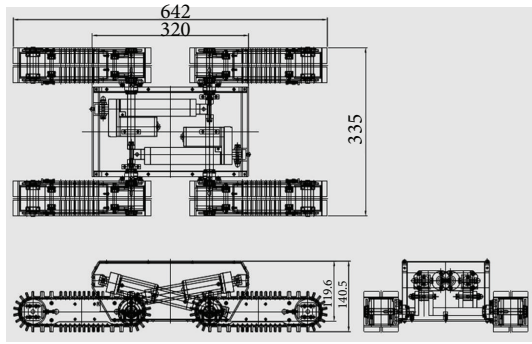


Figure 2: Dimensions of the NT-Hazard Escape-1.

determined based on experimental analysis. In [10], a rough-terrain control (RTC) methodology was presented in multi-wheeled mobile robot systems to improve ground traction and reduce power consumption. A key element of the method is able to estimate the wheel ground contact angles.

The methods in [7–9] remain limitations when the robot is traveling on the different ground surfaces with different and longer trajectory shapes. In this paper, we propose a new dual estimation algorithm for the robot's position and wheel slip estimations based on the Kalman filtering technique. First, the EKF is utilized to estimate the robot's positions, velocities and orientation angles, which are used for the feedback control signals. Second, a discrete KF is designed to estimate the slip parameters of the left and right track wheels, which are used to compensate for the velocity constraints and its variance values for the EKF in order to get more accurate estimated values. In the dual estimation algorithm, these two Kalman filters are performed simultaneously based on the measurements from AHRS sensor and two incremental encoders. The experimental results show that the advantages of the new method overcome the slip limitations in the path tracking control problem of a 4-TW SSMR. The experiment is performed on the NT-Hazard Escape-1, as shown in Figure 1.

This paper is organized as follows: Section 2 introduces the experimental setup and implemented control system. Dual estimation algorithm is presented in Section 3. The experimental results, including analysis and evaluation, are discussed in Section 4. Finally, Section 5 presents the conclusion.

2. The Experimental Setup and Implemented Control System

2.1. The Experimental Setup. The NT-Hazard Escape-1 mobile robot, with four NT-track wheels, considered in this study is applied in the investigation of disaster areas, running and watching rough areas including construction site and running up stairways. The robot with small actuators was made for the first time in Korea, providing various functions to robots. Each of the eight motors is faithful to its own move, and the robot can easily lift two tracks even when is fully loaded. The robot can also move while maintaining its central balance, like a tank. In addition, RS232C is supported to enable working with an embedded board and computer through wireless telecommunication. Therefore, it is easily applied in various applications. The dimensions for the mobile robot are shown in Figure 2.

For research purposes, an experimental mobile robot was designed. The system configuration was represented by a configuration for all of the elemental components as shown in Figure 3(a), while the real photograph of the experimental system is displayed in Figure 3(b). From these figures, the AHRS (AHRS-03-300) was mounted on the top and center of the robot based on a base platform. Two incremental encoders (E40H-8-1024-3-N-5) were installed on the robot's front-left and right tracked wheels. A notebook (model ASUS U36S) was used to collect the data from the AHRS sensor and two incremental encoders via an AVR 128-pro board. This notebook also performed all algorithms which were applied in real-time. Here, the control algorithm for the system was coded in C# programming language (using the Microsoft Visual Studio 2008 version).

During the experiment, the measurement sampling frequency for the AHRS and incremental encoders is set to 20 Hz. The robot is programmed to follow a desired trajectory. Based on the measurements from these two sensors, the robot position values are derived through the Kalman filtering technique, and then robot controllers are applied to drive the robot to the desired trajectories.

2.2. The Implemented Control System. In the experimental mobile robot, a path tracking control method for the robot motion is performed to follow the desired trajectory based on closed-loop control systems and the sensor fusion technique for feedback control signals. Figure 4 describes the diagram of the implemented control algorithm. As shown in this figure, the control system includes three closed-loop controls: two low-level closed-loop controls for motor speeds of the left and right sides and one high-level closed-loop control for robot position. The discrete PID controllers are applied for low-level closed-loop controls. For the path tracking control system, the stable tracking control (STC) method is introduced in [11, 12], based on an error model of the kinematical model. In another approach, vector field orientation (VFO) feedback control method is a motivating control technique to calculate such linear and angular velocities for DDVs. The

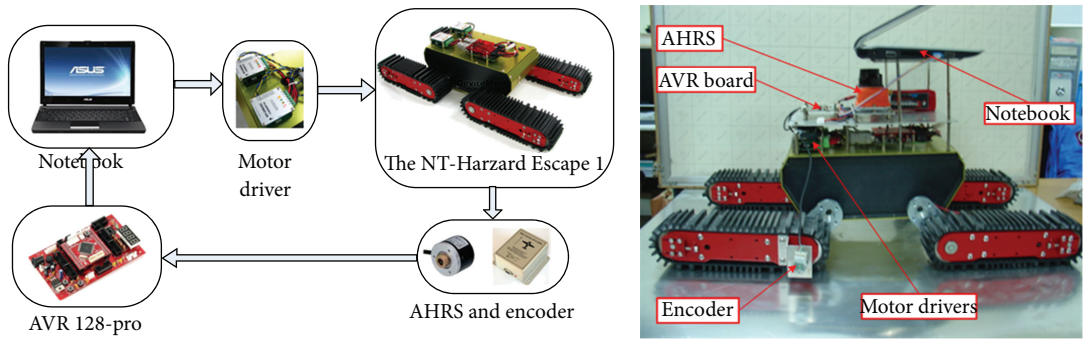


Figure 3: (a) The system configuration. (b) The real photograph of the experimental system.

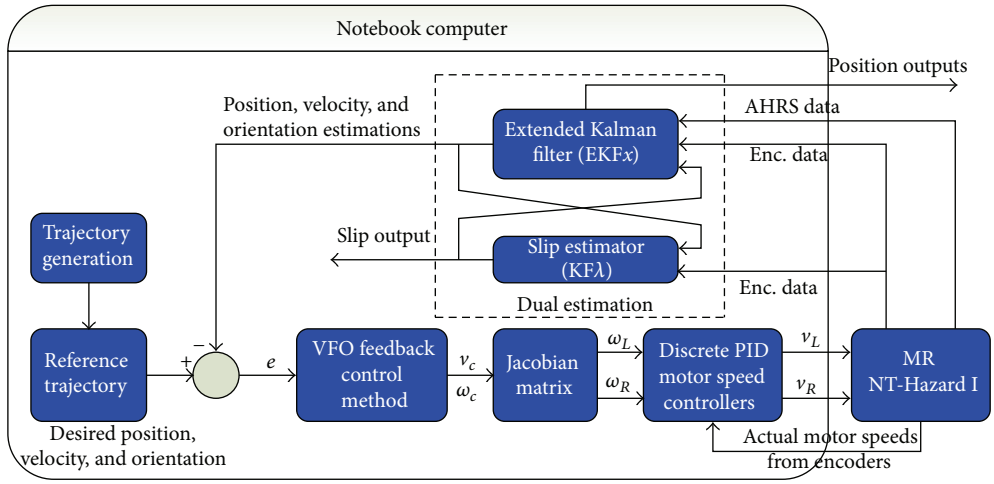


Figure 4: The diagram of the implemented control system.

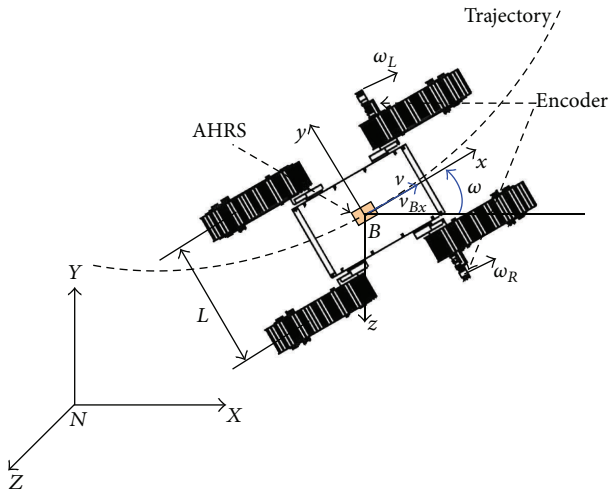


Figure 5: An AHRS kinematics model of a 4-TW SSMR.

reader can refer to [13] (an application for a Four wheel drive skid-steered mobile robot), [14] (VFO for a DDV) and [15] (an extension of VFO in the case of skid-slip phenomena). In this work, the VFO control feedback method is applied for the high-level closed-loop control of the robot position in a 2D plane.

Additionally, the dual estimation algorithm is proposed by integrating the EKF and the discrete KF. First, the EKF, denoted by EKFx, is utilized to estimate the positions, velocities and orientation angles which are used for the feedback control signals in the high-level closed-loop control. Second, a discrete KF, denoted by KFλ, is designed to estimate the slip parameters of the left and right track wheels, which are used to compensate the velocity constraints and its variance values for the EKFx in order to get more accurate estimation values. In the dual estimation algorithm, these EKFx and KFλ are performed simultaneously based on the measurements from AHRS sensor and two incremental encoders, and they are introduced in the next sections.

3. Dual Estimation Algorithm

3.1. AHRS Kinematics and Velocity Constraints. We define a navigational reference frame $N(X, Y, Z)$ and robot body frame $B(x, y, z)$ as shown in Figure 5. Let $P_N(t) = [X_N(t), Y_N(t), Z_N(t)]^T \in \mathbb{R}^3$, $V_N(t) = [V_x(t), V_y(t), V_z(t)]^T \in \mathbb{R}^3$, and $\Theta(t) = [\phi(t), \theta(t), \psi(t)]^T \in \mathbb{R}^3$ denote the position,

velocity, and attitude angle vectors of AHRS in N frame, respectively. We also define the AHRS acceleration, angular rate, acceleration walking bias, and angular rate walking bias in the B frame as $a_B = [a_{Bx}, a_{By}, a_{Bz}]^T \in \mathbb{R}^3$, $\omega_B = [\omega_{Bx}, \omega_{By}, \omega_{Bz}]^T \in \mathbb{R}^3$, $b_{aB} = [b_{aBx}, b_{aBy}, b_{aBz}]^T \in \mathbb{R}^3$, and $b_{gB} = [b_{gBx}, b_{gBy}, b_{gBz}]^T \in \mathbb{R}^3$, respectively.

After subtracting the constant offset and local gravity vector, the acceleration and angular rate models of the AHRS can be described by [16]

$$\begin{aligned} a_B &= \ddot{x}_B + b_{aB} + w_{\text{accel}}, \\ \omega_B &= r_B + b_{gB} + w_{\text{gyros}}, \end{aligned} \tag{1}$$

with

$$\begin{aligned} \dot{b}_{aB} &= -\frac{1}{\tau_a} b_{aB} + \sqrt{\frac{2 f_s \sigma_{baB}^2}{\tau_a}}\, w_{\text{accbias}}, \\ \dot{b}_{gB} &= -\frac{1}{\tau_g} b_{gB} + \sqrt{\frac{2 f_s \sigma_{bgB}^2}{\tau_g}}\, w_{\text{gyrobias}}, \end{aligned} \tag{2}$$

where the true acceleration vector, true angular rate vector, acceleration white noise, and angular rate white noise are $\ddot{x}_B = [\ddot{x}_{Bx}, \ddot{x}_{By}, \ddot{x}_{Bz}]^T$, $r_B = [r_{Bx}, r_{By}, r_{Bz}]^T$, w_{accel}, and w_{gyros}, respectively. And τ_a, σ_{baB}^2, w_{accbias}, τ_g, σ_{bgB}^2 and w_{gyrobias} are time constants, noise variances, and zero-mean white noises with $E[w^2] = 1$ of the acceleration walking bias and angular rate walking bias, and the f_s is the sampling frequency.

We also define the state vector $\chi(t) = [P_N(t), V_N(t), \Theta(t), b_{aB}(t), b_{gB}(t)] \in \mathbb{R}^{15}$ of model process. The kinematic motion equation for AHRS can be simplified to:

$$\dot{\chi} = \begin{bmatrix} V_I \\ C_B^N \left(a_B - b_{aB} + w_{\text{accel}}\right) \\ q_\Theta \left(\omega_B - b_{gB} + w_{\text{gyros}}\right) \\ -\dfrac{1}{\tau_a} b_{aB} \\ -\dfrac{1}{\tau_g} b_{gB} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \sqrt{\dfrac{2 f_s \sigma_{baB}^2}{\tau_a}}\, w_{\text{accbias}} \\ \sqrt{\dfrac{2 f_s \sigma_{bgB}^2}{\tau_g}}\, w_{\text{gyrobias}} \end{bmatrix}, \tag{3}$$

where C_B^N and q_Θ, the transformation matrix from the B frame to the N frame and transformation matrix of the Euler angles, as given by the following matrices:

$$C_B^N = \begin{bmatrix} c_\theta c_\psi & -s_\psi c_\phi + c_\psi s_\phi s_\theta & s_\phi s_\psi + c_\psi s_\theta c_\phi \\ c_\theta s_\psi & c_\phi c_\psi + s_\theta s_\phi s_\psi & -s_\phi c_\psi + s_\theta c_\phi s_\psi \\ -s_\theta & c_\theta s_\phi & c_\phi c_\theta \end{bmatrix}, \tag{4}$$

$$q_\Theta = \begin{bmatrix} 1 & s_\phi t_\theta & c_\phi t_\theta \\ 0 & c_\theta & -s_\theta \\ 0 & \dfrac{s_\phi}{c_\theta} & \dfrac{c_\phi}{c_\theta} \end{bmatrix}, \tag{5}$$

where $c_\theta = \cos\theta$, $s_\theta = \sin\theta$, $t_\theta = \tan\theta$, and the same notation convention is used for angles ϕ and ψ.

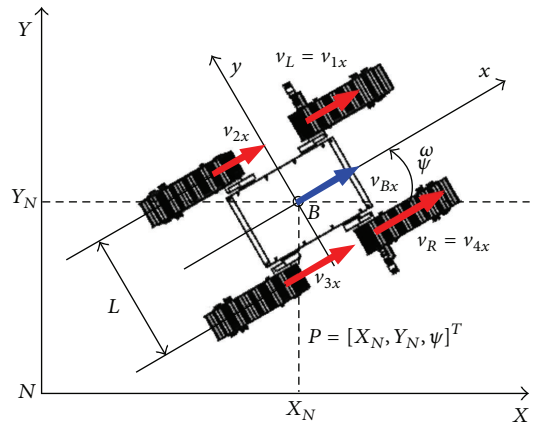


Figure 6: The 4-TW SSMR kinematics.

In the 4-TW SSMR, the motors that power the wheels at each side are geared internally to ensure that the velocities of two adjacent wheels at each side are synchronized (have the same angular velocity) and thus have the same velocity at ground contact. Let us define v_L and v_R as the left and right linear velocities of the robot; we have

$$\begin{aligned} v_L &= v_{1x} = v_{2x}, \\ v_R &= v_{3x} = v_{4x}, \end{aligned} \tag{6}$$

where v_{1x}, v_{2x}, v_{3x}, and v_{4x} are center linear velocities for front-left, rear-left, front-right, and rear-right wheels, respectively, shown in Figure 6. The longitudinal wheel slips of the left and right wheels λ_L and λ_R are defined as ratios of the wheel velocities and its center velocities [17], as follows:

$$\begin{aligned} \lambda_L &= \frac{r\omega_L - v_L}{r\omega_L}, \\ \lambda_R &= \frac{r\omega_R - v_R}{r\omega_R}, \end{aligned} \tag{7}$$

where r is the wheel radius, ω_L and ω_R are the wheel angular speeds for the left and right sides of the mobile robot. From (7), it can be seen that the wheel slip $\lambda \in [0, 1]$ if the wheel is under traction and $\lambda \in (-\infty, 0]$ if the wheel is under braking. Using the definition of the slip in (7), we have

$$\begin{aligned} v_L &= (1 - \lambda_L)\, r\omega_L, \\ v_R &= (1 - \lambda_R)\, r\omega_R. \end{aligned} \tag{8}$$

Because of the symmetric mechanical structure of the robot, it can be assumed that the center of mass (COM) of the robot is located at the center of geometry (COG) of the body frame. The AHRS coordinate is located at the COG in B frame, as shown in Figure 5. Using two wheel encoders, we obtain the AHRS velocity vector $v_B = [v_{Bx}, v_{By}, v_{Bz}]^T$ in B frame. From (8), using two wheel encoder measurements and wheel slip parameters, we obtain the longitudinal velocity v_{Bx} as follows:

$$v_{Bx} = \frac{r}{2}\left[(1 - \lambda_L)\,\omega_L + (1 - \lambda_R)\,\omega_R\right]. \tag{9}$$

Based on [17], since the four tracked wheels of our robot are always in contact with the ground and since the AHRS is fixed on the robot platform, the velocity constraints v_{By} and v_{Bz} in y-axis and z-axis directions for the AHRS device can be simplified to equal zero.

The noises in the longitudinal velocity v_{Bx} can be expressed as the sum of the noises of the left and right wheels' angular speeds ω_L and ω_R as measured by encoders. It is assumed that the noises for ω_L and ω_R have a normal distribution with a zero-mean and corresponding variances, and there is no cross-correlation between the noise of ω_R and of ω_L [18]. The noise variance of the longitudinal velocity v_{Bx} can be expressed as

$$\sigma_{vBx}^2 = \frac{r^2}{4}\left(\left(1-\lambda_R\right)^2 \mathrm{var}\left(\omega_R\right) + \left(1-\lambda_L\right)^2 \mathrm{var}\left(\omega_L\right)\right). \quad (10)$$

The noise variances σ_{vBy}^2 and σ_{vBz}^2 of the lateral velocity v_{By} and ground surface topography v_{Bz} are also expressed by zeros. In order to obtain the longitudinal velocity v_{Bx} and its noise variance, we have to estimate slips λ_L and λ_R. The proposed dual estimation method, which can estimate these two slip values, is introduced in Section 3.2.

3.2. Dual Estimation Algorithm. Now, we define the state variable's vector $X(k) = [P_N(k), V_N(k), \Theta(k), b_{aB}(k), b_{gB}(k)]^T \in \mathbb{R}^{15}$. We rewrite the AHRS kinematics equation shown in (3) in discrete-time form, and we obtain the process model as

$$\begin{aligned} X(k) &= f\left(X(k-1), u(k), w(k), \Delta T\right) \\ &= X(k-1) + \Delta T * g\left(X(k-1), u(k), w(k)\right), \end{aligned} \quad (11)$$

where the AHRS input signals at the kth sampling time is $u(k) = [a_B(k), \omega_B(k)]^T \in \mathbb{R}^6$ and the data sampling period is ΔT. Also, the nonlinear function g of the process model is

$$\begin{aligned} & g\left(X(k-1), u(k), w(k)\right) \\ &= \begin{bmatrix} g_P \\ g_V \\ g_\Theta \\ g_{aB} \\ g_{gB} \end{bmatrix} \\ &= \begin{bmatrix} V_I(k-1) \\ C_B^N(k-1)\left(a_B(k) - b_{aB}(k-1) + w_{\text{accel}}(k)\right) \\ q_\Theta(k-1)\left(\omega_B(k) - b_{gB}(k-1) + w_{\text{gyros}}(k)\right) \\ -\dfrac{1}{\tau_a} b_{aB}(k-1) \\ -\dfrac{1}{\tau_g} b_{gB}(k-1) \end{bmatrix} \\ &+ \begin{bmatrix} 0 \\ 0 \\ 0 \\ \sqrt{\dfrac{2 f_s \sigma_{baB}^2}{\tau_a}}\, w_{\text{accbias}}(k) \\ \sqrt{\dfrac{2 f_s \sigma_{bgB}^2}{\tau_g}}\, w_{\text{gyrobias}}(k) \end{bmatrix}, \end{aligned} \quad (12)$$

and $w(k) = [w_{\text{accle}}(k), w_{\text{gyros}}(k), w_{\text{accbias}}(k), w_{\text{gyrobias}}(k)]^T \in \mathbb{R}^{12}$ is the process noise vector.

With velocity constraints (9), the AHRS velocities in B frame are considered as the measurement vector $y(k) = [v_{Bx}(k), v_{By}(k), v_{Bz}(k)]^T \in \mathbb{R}^3$. Including the wheel encoder measurement noise and ground topography, the measurement model is rewritten in discrete-time form as

$$y(k) = h\left(X(k) + n(k)\right) = \left(C_B^N\right)^T V_N + n(k), \quad (13)$$

where $n(k)$ is measurement noises. We assume that the measurement noises $n(k)$ are independent, zero-mean, and Gaussian white noise; that is, $n(k) \sim N(0, R)$. The covariance matrix of the measurement noise is

$$R(k) = \mathrm{diag}\left(\sigma_{vBx}^2, \sigma_{vBy}^2, \sigma_{vBz}^2\right) \in \mathbb{R}^{3\times 3}, \quad (14)$$

where σ_{vBx}^2, σ_{vBy}^2, and σ_{vBz}^2 are calculated in (10). The EKFx was implemented by using systems (11) and (13) in order to obtain the estimated positions, velocities, and attitudes of the robot.

Additionally, in order to determine wheel slip parameters to compensate the velocity constraints and its variance values, a discrete KF, denoted by KFλ, is designed. By giving v_{Bx} in body frame B and yaw rate $\dot{\psi}$, the wheel center velocities along x-axis for the left and right sides wheels are calculated by

$$\begin{aligned} v_L &= v_{Bx} - \frac{L}{2}\dot{\psi}, \\ v_R &= v_{Bx} + \frac{L}{2}\dot{\psi}, \end{aligned} \quad (15)$$

where L is lateral wheel bases. By substituting (15) into (7), we have

$$\begin{aligned} \lambda_L &= \frac{r\omega_L - v_{Bx} + (L/2)\,\dot{\psi}}{r\omega_L}, \\ \lambda_R &= \frac{r\omega_R - v_{Bx} - (L/2)\,\dot{\psi}}{r\omega_R}, \end{aligned} \quad (16)$$

$$\begin{aligned} \Longrightarrow \quad \omega_L &= \frac{1}{1-\lambda_L}\left[\frac{v_{Bx}}{r} - \frac{L\dot{\psi}}{2r}\right], \\ \omega_R &= \frac{1}{1-\lambda_R}\left[\frac{v_{Bx}}{r} + \frac{L\dot{\psi}}{2r}\right]. \end{aligned} \quad (17)$$

We define unknown parameters as:

$$\begin{aligned} \gamma_L &= \frac{1}{1-\lambda_L}; \qquad \gamma_R = \frac{1}{1-\lambda_R}, \\ \lambda &\in (-\infty, 1] \Longrightarrow \gamma \in [0, +\infty). \end{aligned} \quad (18)$$

Substituting (18) into (17), the relationship between the angular speed from encoders and unknown parameters is described as

$$\begin{bmatrix}\omega_L\\ \omega_R\end{bmatrix} = \begin{bmatrix}\frac{v_{Bx}}{r} - \frac{L\dot{\psi}}{2r} & 0\\ 0 & \frac{v_{Bx}}{r} + \frac{L\dot{\psi}}{2r}\end{bmatrix}\begin{bmatrix}\gamma_L\\ \gamma_R\end{bmatrix}. \tag{19}$$

The slip estimation based on the KFλ is now established. Assuming that the slip process is the random process, then the unknown parameter process can be modeled in discrete-time form as

$$\gamma(k) = \Phi_k \gamma(k-1) + w_\gamma(k-1), \tag{20}$$

where $\gamma(k) = [\gamma_L(k), \gamma_R(k)]^T$ is the process state vector of the KFλ at time k step, $w_\gamma(k) = [w_{\gamma 1}(k), w_{\gamma 2}(k)]^T$ is process noise vector, and Φ_k is the state transition matrix which is defined as:

$$\Phi_k = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}. \tag{21}$$

From (19), the measurement model of the KFλ can be written relating to the encoder measurement noises as

$$z(k) = H_\gamma(k)\gamma(k) + n_\gamma(k), \tag{22}$$

where $z(k) = [\omega_L(k), \omega_R(k)]^T$ is angular speed vector from two encoders at time k step, $n_\gamma(k) = [n_{\gamma 1}(k), n_{\gamma 2}(k)]^T$ is measurement noise vector, and $H_\gamma(k)$ is the measurement matrix which is calculated as follows:

$$H_\gamma(k) = \begin{bmatrix}\left(\frac{\widehat{v}_{Bx}(k)}{r} - \frac{L\widehat{\dot{\psi}}(k)}{2r}\right) & 0\\ 0 & \left(\frac{\widehat{v}_{Bx}(k)}{r} + \frac{L\widehat{\dot{\psi}}(k)}{2r}\right)\end{bmatrix}, \tag{23}$$

where $\widehat{v}_{Bx}(k)$ and $\widehat{\dot{\psi}}(k)$ are the estimated values of the linear and angular velocities of mobile robot in B frame which are computed as

$$\begin{aligned}\widehat{v}_{Bx}(k) &= \widehat{V}_x(k)\cos\left(\widehat{\psi}(k)\right) + \widehat{V}_y(k)\sin\left(\widehat{\psi}(k)\right),\\ \widehat{\dot{\psi}}(k) &= \frac{\widehat{\psi}(k) - \widehat{\psi}(k-1)}{\Delta T},\end{aligned} \tag{24}$$

where $\widehat{V}_x(k)$, $\widehat{V}_y(k)$, and $\widehat{\psi}(k)$ are given from the EKFx. The Systems (20) and (22) are used to perform the KFλ in order to estimate the unknown parameters of the left and right sides. From (18), the estimated slip parameters $\widehat{\lambda}_L(k)$ and $\widehat{\lambda}_R(k)$ at k step are calculated through the estimated unknown parameters $\widehat{\gamma}_L(k)$ and $\widehat{\gamma}_R(k)$ as follows:

$$\begin{aligned}\widehat{\lambda}_L(k) &= 1 - \frac{1}{\widehat{\gamma}_L(k)},\\ \widehat{\lambda}_R(k) &= 1 - \frac{1}{\widehat{\gamma}_R(k)}.\end{aligned} \tag{25}$$

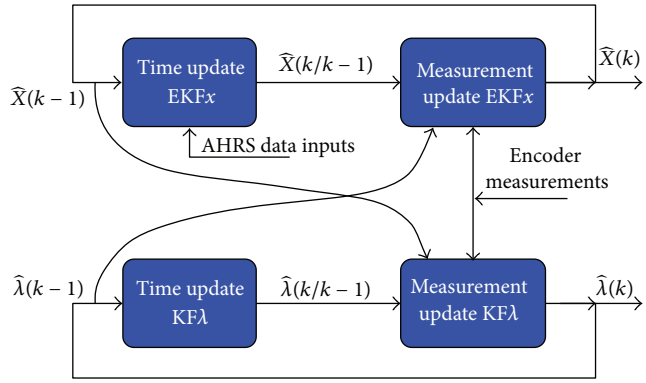


Figure 7: The dual estimation.

From previous descriptions, the estimated slip values $\widehat{\lambda}_L$ and $\widehat{\lambda}_R$ are obtained indirectly through the unknown parameters $\widehat{\gamma}_L$ and $\widehat{\gamma}_R$. In this method, the current values $\widehat{v}_{Bx}(k)$ and $\widehat{\dot{\psi}}(k)$ (in (24)) are calculated by using the current state $\widehat{X}(k)$ and $\widehat{X}(k-1)$ of the EKFx. Similarly, the current estimated slip values in (25) are used to compute the longitudinal velocity $v_{Bx}(k)$ in (9) and its noise variance value in (10) in order to perform the EKFx. From the previous analyses, in this work, there are two Kalman filters running concurrently; it is called *the dual estimation algorithm* [19, 20]. As illustrated in Figure 7, the algorithm consists of two kinds of Kalman filters that run concurrently. The top EKFx generates state estimates and requires $\widehat{\lambda}(k-1)$ for the measurement update. The bottom KFλ generates slip state estimates and requires $\widehat{X}(k-1)$ for the measurement update. The following describes the dual estimation process through mathematical equations [21, 22].

Initialize with

$$\begin{aligned}\widehat{\gamma}(0) &= E[\gamma(0)],\\ P_\gamma(0) &= E\left[(\gamma(0) - \widehat{\gamma}(0))(\gamma(0) - \widehat{\gamma}(0))^T\right],\\ \widehat{X}(0) &= E[X(0)],\\ P(0) &= E\left[\left(X(0) - \widehat{X}(0)\right)\left(X(0) - \widehat{X}(0)\right)^T\right].\end{aligned} \tag{26}$$

For $k \in \{1, \ldots, \infty\}$, the prediction step for KFλ is

$$\begin{aligned}\widehat{\gamma}(k/k-1) &= \Phi_{k-1}\widehat{\gamma}(k-1)\\ P_\gamma(k/k-1) &= \Phi_{k-1}P_\gamma(k-1)\Phi_{k-1}^T + Q_\gamma,\end{aligned} \tag{27}$$

and update step equations for KFλ are

$$H_\gamma(k) = \begin{bmatrix} \left(\frac{\hat{v}_{Bx}(k-1)}{r} - \frac{L\hat{\dot{\psi}}(k-1)}{2r}\right) & 0 \\ 0 & \left(\frac{\hat{v}_{Bx}(k-1)}{r} + \frac{L\hat{\dot{\psi}}(k-1)}{2r}\right) \end{bmatrix},$$

$$K_\gamma(k) = P_\gamma(k/k-1)\,H_\gamma^T(k) \times \left[H_\gamma(k)P_\gamma(k/k-1)H_\gamma^T(k) + R_\gamma(k)\right]^{-1}$$

$$\hat{\gamma}(k) = \hat{\gamma}(k/k-1) + K_\gamma(k)\left[z(k) - H_\gamma(k)\,\hat{\gamma}(k/k-1)\right],$$

$$P_\gamma(k) = \left[I - K_\gamma(k)\,H_\gamma(k)\right]P_\gamma(k/k-1)$$

$$\hat{\lambda}_L(k) = 1 - \frac{1}{\hat{\gamma}_L(k)}$$

$$\hat{\lambda}_R(k) = 1 - \frac{1}{\hat{\gamma}_R(k)}, \tag{28}$$

where Q_γ and $R_\gamma(k)$ are the process and measurement covariance matrices of $w_\gamma(k)$ and $n_\gamma(k)$, respectively.

And the prediction step for EKFx is

$$\hat{X}(k/k-1) = \hat{X}(k-1) + \Delta T g\left(\hat{X}(k-1), u(k), 0\right),$$

$$P(k/k-1) = F_X(k)\,P(k-1)\,F_X^T(k) + W(k)\,QW^T(k), \tag{29}$$

and update step for EKFx is

$$v_{Bx}(k) = \frac{r}{2}\left[\left(1-\hat{\lambda}_L(k-1)\right)\omega_L(k) + \left(1-\hat{\lambda}_R(k-1)\right)\omega_R(k)\right]$$

$$\sigma_{vBx}^2 = \frac{r^2}{4}\left(\left(1-\hat{\lambda}_R(k-1)\right)^2 \mathrm{var}(\omega_R) + \left(1-\hat{\lambda}_L(k-1)\right)^2 \mathrm{var}(\omega_L)\right),$$

$$S(k) = H_X(k)\,P(k/k-1)\,H_X^T(k) + R(k),$$

$$K(k) = P(k/k-1)\,H_X^T(k)\,S^{-1}(k),$$

$$\hat{X}(k) = \hat{X}(k/k-1) + K(k)\left[y(k) - h\left(\hat{X}(k/k-1), 0\right)\right],$$

$$P(k) = P(k/k-1) - K(k)\,H_X(k)\,P(k/k-1), \tag{30}$$

where

$$F_X(k) = \left.\frac{\partial f(X, u(k), 0)}{\partial X}\right|_{\hat{X}(k-1)},$$

$$W(k) = \left.\frac{\partial f}{\partial w}\right|_{(\hat{X}(k-1), u(k))},$$

$$H_X(k) = \left.\frac{\partial h(X, 0)}{\partial X}\right|_{\hat{X}(k/k-1)}, \tag{31}$$

and Q is the process covariance matrix of w.

Update estimation values $\hat{v}_{Bx}(k)$ and $\hat{\dot{\psi}}(k)$ for KFλ as

$$\hat{v}_{Bx}(k) = \hat{V}_x(k)\cos\left(\hat{\psi}(k)\right) + \hat{V}_y(k)\sin\left(\hat{\psi}(k)\right),$$

$$\hat{\dot{\psi}}(k) = \frac{\hat{\psi}(k) - \hat{\psi}(k-1)}{\Delta T}. \tag{32}$$

4. Experimental Results

In this section, the experiments are performed to prove the effectiveness of the proposed dual estimation algorithm (DEA) for the robot's path tracking problem in real-time application. The VFO feedback control method is applied for the trajectory tracking control. In fact, the wheel slips, which always occur and vary while the robot is traveling on the different ground surfaces with different trajectory shapes, limit the traction and braking abilities of the robot, especially in 4-TW SSMR. Therefore, a proposed DEA is performed to overcome the previous limitations in this research. The estimated values such as robot's positions, velocities, attitudes and wheel slip parameters are obtained more accurately, which are used for the feedback control signals in the high-level closed-loop control. In this work, the experiments have been carried out on concrete terrain in an outdoor environment, involving two scenarios.

Scenario 1. The mobile robot is controlled to follow a desired trajectory with two curves.

Scenario 2. The mobile robot is controlled to follow a desired circular trajectory with a 2 m radius.

In the first scenario, the output trajectories are shown in Figure 9(a). As illustrated in Figure 9(a), the output response without DEA is quite far to the desired response while that with DEA is close to the desired response. The linear and angular velocities of two methods (without and with DEA), shown in Figures 8(a) and 8(b), are also close to the desired velocities despite the mechanical vibrations and sensor noises. In fact, although the working environment contains large numbers of nonlinear and uncertain factors, which impact with the system performances, the results point out that the proposed DEA has improved the performance of the path tracking problem. The previous evaluation is proved by considering the position errors as shown in Figure 9(b). In Figure 9(b), the position error with DEA at any instant of time, which is the distance from the estimated position of the robot at the defined instant to the nearest point of the desired trajectory, is the closest to zero compared to that without DEA. Consequently, the root mean square error (RMSE) of robot position with DEA is the smallest (0.0239 [m]), while that without DEA is 0.0845 [m].

In order to evaluate the proposed method, the system is also tested with a circular trajectory, in the second scenario. The output responses without and with DEA are shown in Figure 11(a). It can see that the response with DEA is still the closest to the desired response. The linear and angular velocities are shown in Figures 10(a) and 10(b). Figure 11(b)

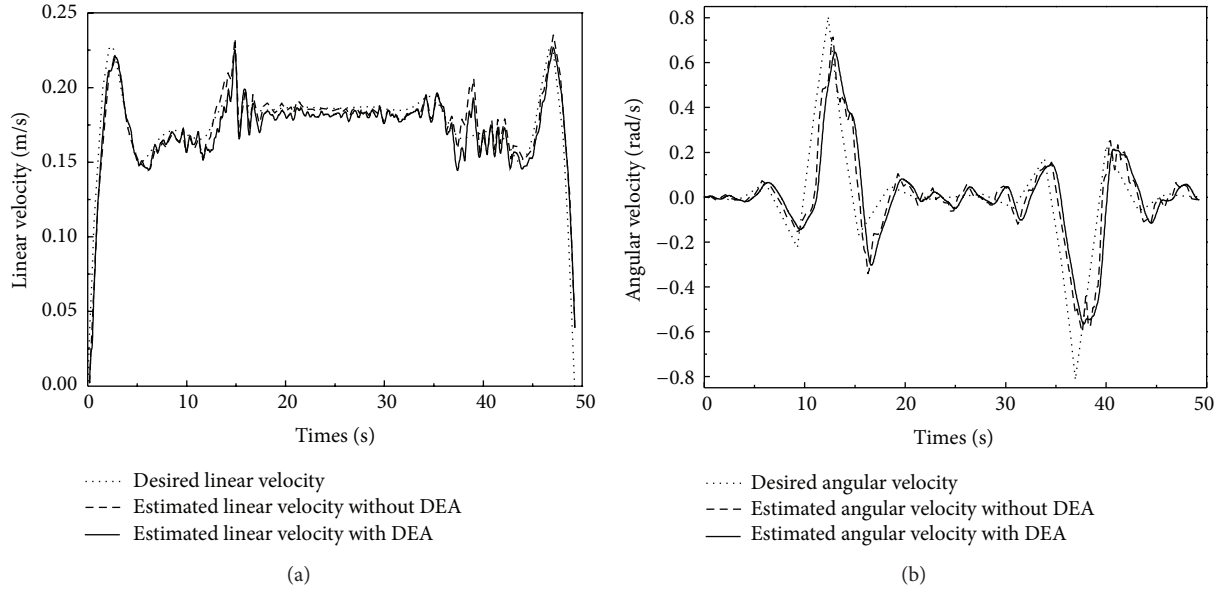


Figure 8: (a) The desired and estimated linear velocities without and with DEA in scenario 1. (b) The desired and estimated angular velocities without and with DEA in scenario 1.

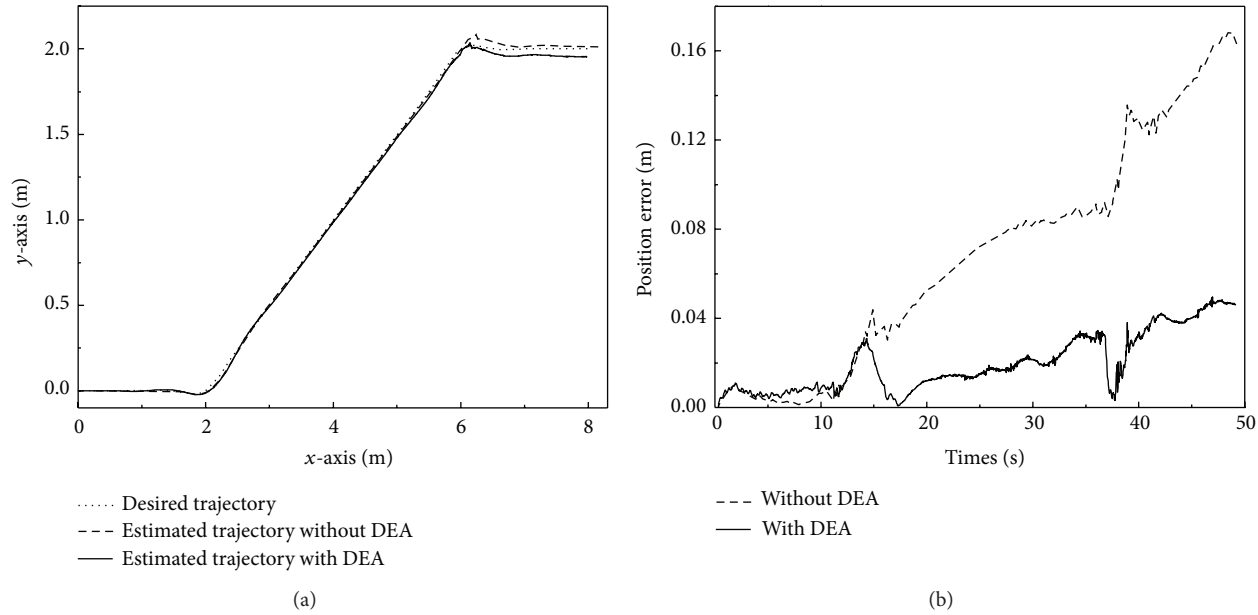


Figure 9: (a) The desired trajectory, estimated trajectories without and with DEA in scenario 1. (b) The position errors without and with DEA in scenario 1.

clearly depicts that the result with DEA is the best while comparing to that without DEA; the position error graph with DEA is still the closest to zero. Hence, the RMSE of robot position with DEA is the smallest (0.0236 [m]), while that without DEA is 0.0882 [m].

In both scenarios, the results show that the trajectories with DEA are the closest to the desired responses. It results from the estimated slip values which are used to compensate the velocity constraints and its variances. The estimated values in EKFx such as velocities and trajectories with DEA are more precise than those without DEA. It means that the feedback control signals with DEA are more accurate, and then the high-level closed-loop control with DEA performs better than the previous testbed without DEA.

From the previous analyses, the robot positioning results with DEA are better than the previous results without DEA. It is clear that the DEA, with wheel slip compensation, is very essential in the mobile robot field. For the first scenario,

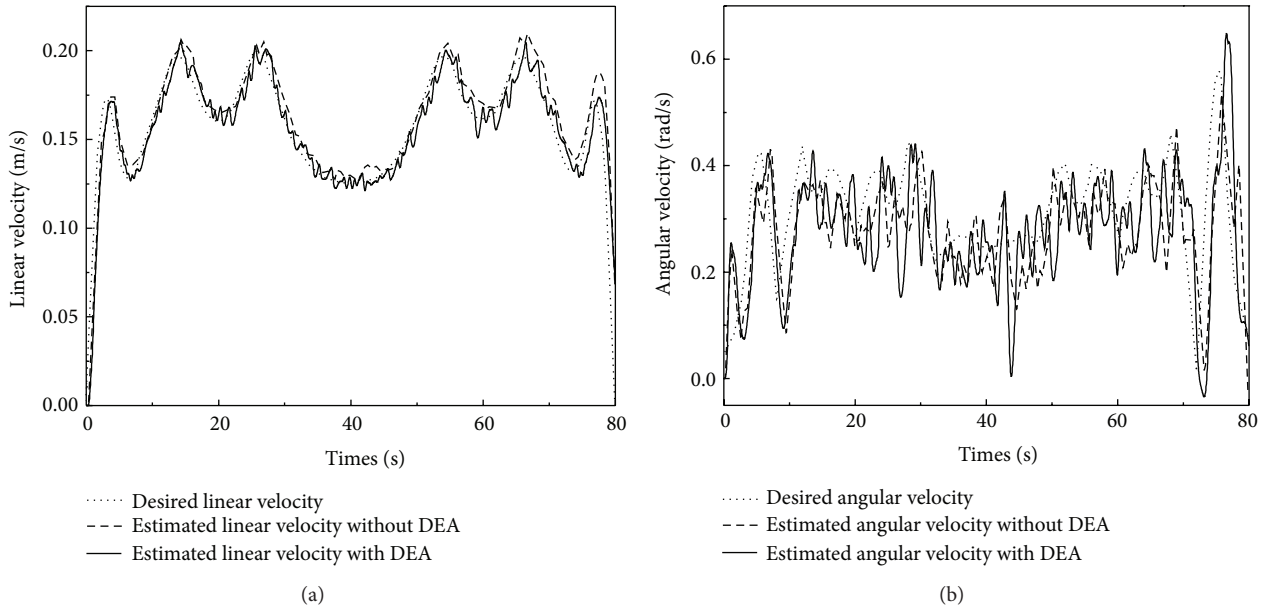


Figure 10: (a) The desired and estimated linear velocities in scenario 2. (b) The desired and estimated angular velocities in scenario 2.

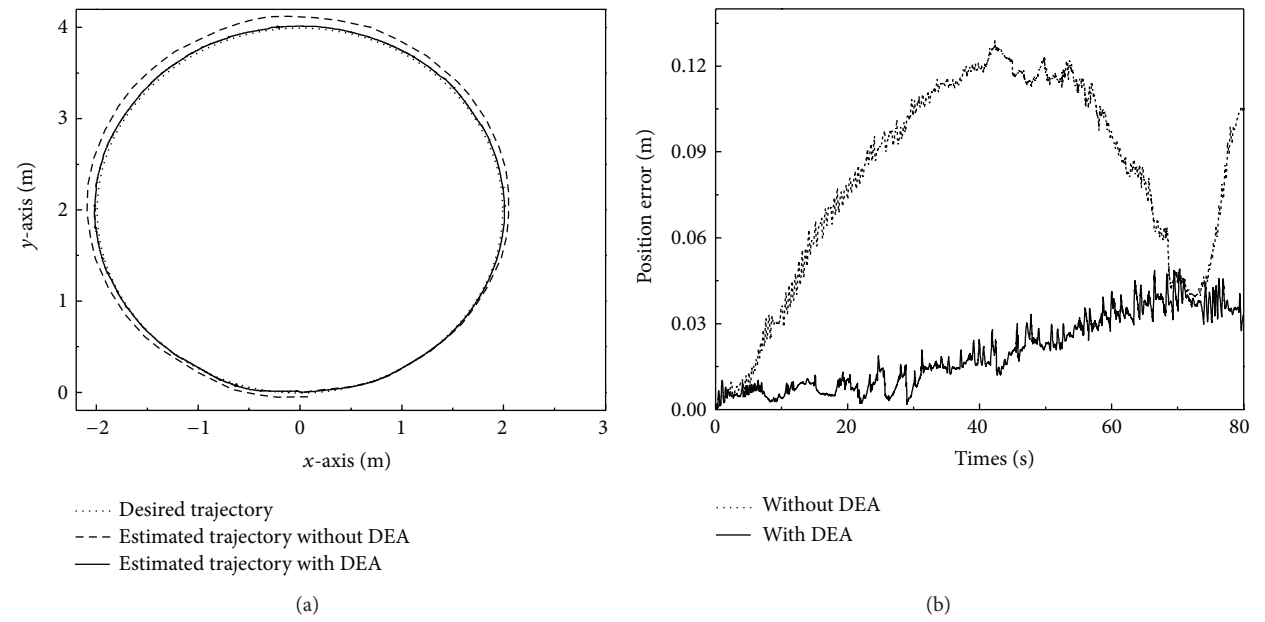


Figure 11: (a) The desired trajectory and estimated trajectories without and with DEA in scenario 2. (b) The position errors without and with DEA in scenario 2.

the robot follows the desired trajectory, which has two curves. At the left rotation in the first curve, the angular speed of the right wheels is greater than that of the left wheels. Then, the traction phenomenon occurs on the right side, while the braking phenomenon appears on the left. Consequently, the slip value of the right wheels is positive, and that of the left wheels is negative. Similarly, at the right rotation in the second curve, the slip value of the right wheels is negative while that of the left side is positive. Furthermore, when the robot travels along the straight segment in the first trajectory, less force is needed to drive the robot and the left and right wheel slips are close to zero. Figure 12(a) illustrates the estimated slip parameters of the left and right wheels in the first scenario.

For the second scenario, the robot follows the desired circular trajectory with counterclockwise direction. The angular

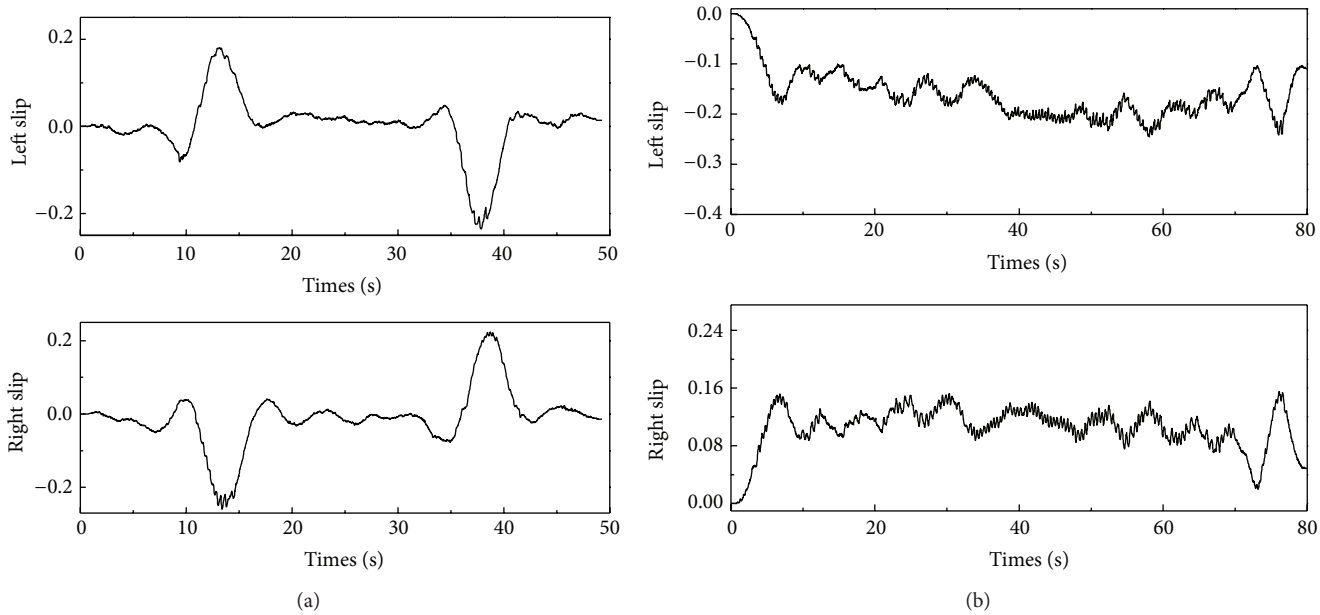


Figure 12: The estimated slip parameters of the left and right wheels in scenarios 1(a) and 2(b).

Table 1: RMSEs of all experiments.

Method	RMSE [m]	
	Scenario 1	Scenario 2
Without DEA	0.0845	0.0882
With DEA	0.0239	0.0236

speed of the right wheels is always larger than that of the left wheels. Consequently, the slip value of the right wheels is always positive and that of the left wheels is always negative, as shown in Figure 12(b). All types of the estimated slip values in both scenarios are consistent. On the whole, the robot positioning results are improved significantly by considering the new dual estimation method. Table 1 shows RMSEs of all experiments in this work; the RMSE values with DEA are reduced more significantly.

5. Conclusion

This paper presents experiments to control the 4-TW SSMR based on the estimated feedback control signals. In the mobile robot field, the wheel slip limits the traction and braking ability of the robot, especially the 4-TW SSMR. The wheel slip always occurs while the robot is traveling on the different ground surfaces with different trajectory shapes. In this paper, we propose the new dual estimation algorithm for robot's positions, velocities, attitudes, and wheel slip parameters based on the Kalman filtering technique in order to obtain more accurate estimation values, which are used for the feedback control signals. In the dual estimation algorithm, two Kalman filters performed simultaneously based on the measurements from AHRS sensor and two incremental encoders. Experimental results show the advantages of the novel method which is effective in the path tracking control problem for a 4-TW SSMR, for example, the commercialized Hazard Escape I mobile robot.

Conflict of Interests

This company does not have any relationship with the name of Hazard Escape 1. So there is no competing interest such as financial gain between two sides.

Acknowledgment

This research was supported by Korea Aerospace Research Institute in 2012.

References

[1] A. J. Weinstein and K. L. Moore, "Pose estimation of Ackerman steering vehicles for outdoors autonomous navigation," in *Proceedings of the IEEE International Conference on Industrial Technology (ICIT '10)*, pp. 579–584, Viña del Mar Valparaíso, Chile, March 2010.

[2] F. Espinosa, C. Santos, M. Marrón-Romera, D. Pizarro, F. Valdés, and J. Dongil, "Odometry and laser scanner fusion based on a discrete extended Kalman filter for robotic platooning guidance," *Sensors*, vol. 11, no. 9, pp. 8339–8357, 2011.

[3] L. Tamás, G. Lazea, R. Robotin, C. Marcu, S. Herle, and Z. Szekely, "State estimation based on Kalman filtering techniques in navigation," in *Proceedings of the IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR '08)*, vol. 2, pp. 147–152, Cluj-Napoca, Romania, May 2008.

[4] L. Teslić, G. Klančar, and I. Škrjanc, "Kalman-filtering-based localization of a mobile robot with a LRF in a simulated 2D

environment," in *Proceedings of the IEEE Mediterranean Electrotechnical Conference (MELECON '08)*, pp. 316–322, Ajaccio, France, May 2008.

[5] L. Yenilmez and H. Temeltas, "Autonomous navigation for planetary exploration by a mobile robot," in *Proceedings of the International Conference on Recent Advances in Space Technologies*, pp. 397–402, November 2003.

[6] D. Wettergreen, B. Shamah, P. Paul Tompkins, and W. Whittaker, "Robotic planetary exploration by sun-synchronous navigation," in *Proceedings of the 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS '01)*, June 2001.

[7] S. Zibin, Y. H. Zweiri, L. D. Seneviratne, and K. Althoefer, "Nonlinear observer for slip estimation of skid-steering vehicles," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '06)*, pp. 1499–1504, Orlando, Fla, USA, May 2006.

[8] X. Song, L. D. Seneviratne, K. Althoefer, and Z. Song, "A robust slip estimation method for skid-steered mobile robots," in *Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision (ICARCV '08)*, pp. 279–284, Hanoi Hoan Kiem District, Vietnam, December 2008.

[9] S. A. A. Moosavian and A. Kalantari, "Experimental slip estimation for exact kinematics modeling and control of a Tracked Mobile Robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '08)*, pp. 95–100, Nice, France, September 2008.

[10] K. Iagnemma and S. Dubowsky, "Mobile Robot Rough-Terrain Control (RTC) for planetary exploration," in *Proceedings of the 26th ASME Biennial Mechanisms and Robotics Conference (DETC '00)*, 2000.

[11] Y. Kanayama, Y. Kimura, F. Miyazaki, and T. Noguchi, "A stable tracking control method for an autonomous mobile robot," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 384–389, May 1990.

[12] E. Maalouf, M. Saad, and H. Saliah, "A higher level path tracking controller for a four-wheel differentially steered mobile robot," *Robotics and Autonomous Systems*, vol. 54, no. 1, pp. 23–33, 2006.

[13] S. Arslan and H. Temeltaş, "Robust motion control of a four wheel drive skid-steered mobile robot," in *Proceedings of the 7th International Conference on Electrical and Electronics Engineering (ELECO '11)*, pp. II415–II419, December 2011.

[14] M. Michałek and K. Kozłowski, "Vector-field-orientation feedback control method for a differentially driven vehicle," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 1, pp. 45–65, 2010.

[15] M. M. Michałek, P. Dutkiewicz, M. Kiełczewski, and D. Pazderski, "Vector-field-orientation tracking control for a mobile vehicle disturbed by the skid-slip phenomena," *Journal of Intelligent and Robotic Systems*, vol. 59, no. 3-4, pp. 341–365, 2010.

[16] W. S. Flenniken IV, J. H. Wall, and D. M. Bevly, "Characterization of various IMU error sources and the effect on navigation performance," in *Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS '05)*, pp. 967–978, September 2005.

[17] J. Yi, H. Wang, J. Zhang, D. Song, S. Jayasuriya, and J. Liu, "Kinematic modeling and analysis of skid-steered mobile robots with applications to low-cost inertial-measurement-unit-based motion estimation," *IEEE Transactions on Robotics*, vol. 25, no. 5, pp. 1087–1097, 2009.

[18] L. Teslić, G. Klančar, and I. Škrjanc, "Kalman-filtering-based localization of a mobile robot with a LRF in a simulated 2D environment," in *Proceedings of the IEEE Mediterranean Electrotechnical Conference (MELECON '08)*, pp. 316–322, Ajaccio, France, May 2008.

[19] E. A. Wan, R. V. D. Merwe, and A. T. Nelson, "Dual estimation and the unscented transformation," in *Advances in Neural Information Processing Systems*, vol. 12, pp. 666–672, November 2000.

[20] S. Haykin, *Kalman Filtering and Neural Networks*, Jokn Wiley & Sons, 1st edition, 2001.

[21] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, University of North Carolina at Chapel Hill, 1995.

[22] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, 3rd edition, 1997.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 247676, 7 pages
http://dx.doi.org/10.1155/2013/247676

Research Article

Assessment of the Kinematic Parameters of Distant Noncooperative Objects

Lovro Kuščer and Janez Diaci

Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva 6, 1000 Ljubljana, Slovenia

Correspondence should be addressed to Lovro Kuščer; lovro.kuscer@fs.uni-lj.si

Received 22 March 2013; Accepted 7 October 2013

Academic Editor: Hongxing Wei

The paper presents a method and a system for assessing the kinematic parameters of distant noncooperative objects. An experimental measuring system was developed using a long distance laser rangefinder, a GPS receiver, an electronic inclinometer, and a CCD camera mounted on a motorized pan-tilt unit. During the measurement process, the system first establishes its position and orientation in a global coordinate system. Particle filtering approach based on adaptive template matching is used to track a moving object of interest in the acquired digital image sequence. The tracking and rangefinder data is employed to determine in real time the relative position of the object, thus obtaining its 3D trajectory and velocity. To enable repetitive range measurements, the tracking data is also used to actuate the pan-tilt unit directing the rangefinder towards the object. Experiments are presented which demonstrate the performance of the system for characterization of movement of vehicles and people at kilometer-range distances.

1. Introduction

The characterization of the movement of different noncooperative objects has been the focus of studies in different fields. The development of modern computer vision systems that employ high performance cameras and advanced image processing algorithms has paved the way for new applications in surveillance, traffic monitoring, people tracking, and elsewhere. The advances and availability of various tracking solutions also enabled the study of human behavior in different environments [1, 2].

A commonly used approach is to utilize one or more cameras placed in convenient locations to observe the passing objects of interest [3, 4]. Then, by employing different image processing techniques, the moving objects are detected, discriminated from one another, and tracked through the image sequence. However, depending solely on visual information has certain shortcomings that become apparent especially when lighting is changed and also when dealing with different types of occlusions. Due to these limitations, visual tracking systems are often combined with laser scanners to improve the reliability [5, 6]. Laser scanners provide additional information at low computational complexity and are used to enhance the tracking performance. In certain cases, laser scanners provide sufficient information to perform people counting and tracking without the use of cameras [2, 7].

However, the majority of studies in the fields of people tracking and traffic monitoring perform the movement characterization at relatively small ranges since the measuring range is often limited by the employed laser scanners which are capable of measuring distances of only a few hundred meters. Also, in cases of commonly used single-row laser scanners, the applications are restricted to relatively flat terrain. This limits the applicability in both urban and nonurban environments.

In the presented work, we investigate the possibilities for real-time assessment of the kinematic parameters of noncooperative objects at kilometer-range distances. The term noncooperative refers to the fact that the objects of interest are not marked, prepared, or equipped in any way that would facilitate the measuring process. To achieve this functionality, we combine a long distance laser rangefinder with a custom visual tracking system. This enables the exploitation of efficient tracking algorithms to maintain the orientation of the laser rangefinder's optical axis towards the distant object. In this way, the repeated measurements of

the object's relative position are performed. The presented approach yields the 3D velocity and trajectory information in a global coordinate system.

This paper is organized as follows. In Section 2, we describe the experimental system and our approach to visual tracking based on particle filtering and adaptive template matching. Section 3 presents the experimental work aimed at measuring vehicle and people kinematics in real time. Furthermore, the obtained results are discussed and compared to the reference measurements. Finally, in Section 4 we summarize the main achievements and expose key limitations of the developed experimental system and the employed methodology.

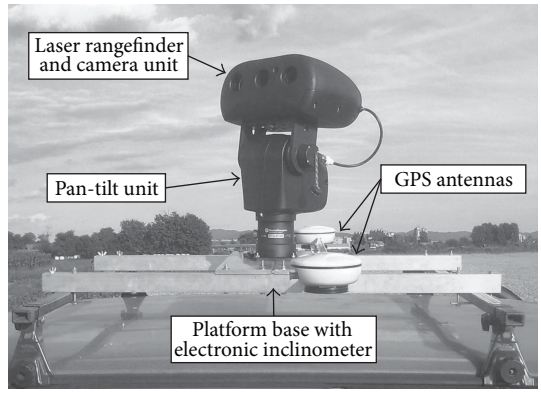


Figure 1: Experimental measuring system.

2. Experimental System

For characterization of kinematic parameters of distant objects, we have developed an experimental measuring system, which is composed of COTS (commercial of-the-shelf) components. It employs an eye-safe laser rangefinder with a measuring range of 50 m to 20 km, a high performance grayscale CCD camera with a 100 mm focal length lens, a GPS receiver/compass, an electronic inclinometer, and an inertially stabilized pan-tilt unit. The camera and the laser rangefinder are mounted on the pan-tilt unit, while the GPS receiver/compass and electronic inclinometer are attached to the platform base (Figure 1). The system is controlled by a personal computer with custom software.

In comparison with a single-row laser scanner, the employed laser rangefinder is capable of measuring longer distances. On the other hand, its measuring frequency is significantly lower since it is designed to measure the distance to a single point, whereas the laser scanner measures the distances to multiple points in its field of view.

The experimental measuring system was initially designed for measuring the geographic location of distant stationary objects [8] and has been subsequently upgraded for measuring the kinematic parameters of moving objects. This was achieved with the integration of a high performance, thermoelectrically cooled CCD camera and appropriate image processing techniques.

During the measurement process, the GPS receiver is used to determine the position of the measuring system with the specified uncertainty of 0.6 m (95% confidence). Next, the relative position of the distant object is measured by the GPS compass, the electronic inclinometer, and the laser rangefinder. According to the manufacturer specification, the measurement uncertainties of the azimuth, elevation, and distance are $\pm 0.3^{\circ}$ (95% confidence), $\pm 0.5^{\circ}$ (95% confidence), and ±3 m (95% confidence), respectively. Using the acquired measurements, the position of the distant object is first determined in the WGS 84 Cartesian coordinate system and then transformed to the WGS 84 geodetic coordinate system utilizing an iterative method [8, 9].

For assessment of the distant object kinematics, the employed CCD camera acquires images at the resolution of 672 × 448 pixels. With the appropriate image processing algorithms, the location of the distant object in the image is determined. This information is then used as an additional measurement and also to actuate the pan-tilt unit in order to maintain the orientation of the laser rangefinder towards the distant object. With the 100 mm focal length lens, the horizontal field of view of the optical system is 2.48° or approximately 0.064 mrad/pixel. This angle is considerably smaller than the 1 mrad beam divergence of the laser rangefinder and thus allows a precise determination of the distant object position.

2.1. Visual Tracking. The task of visual object tracking is to determine the location of the object of interest in every frame of the image sequence. The approach to visual object tracking highly depends on the specifics of each individual application. We conducted a series of preliminary experiments on recorded image sequences of distant moving people and vehicles in order to test various possible approaches and determine a suitable one. We examined the options of relying on color histogram, movement, and intensity distribution information. Based on the results, we selected the adaptive template matching technique [10] as the basis for our method because this approach offers the possibility for at least partial elimination of the effects of changes in lighting and reflections, which often occur in outdoor environment. Additionally, the selected approach depends on the intensity distribution rather than on color information that is less reliable at long distances due to the atmospheric effects.

Our approach to the assessment of the kinematic parameters of distant objects is presented in Figure 2.

The tracking session is started by the operator who selects the object in the live video. A selection is performed with a rectangular region of 30 × 20 pixels which can be resized depending on the size of the tracked object. The selected region serves as an initial template for the template matching procedure. The similarity of a region in current image with the template is assessed with the normalized cross-correlation coefficient (1) which is calculated using a fast implementation in OpenCV library [11]. The utilized correlation approach is

(i) Initialization: particle set $\{x_0^i, 1/N\}_{i=1}^N$ with uniform weights

(ii) For $k = 1, \ldots, K$

(1) Prediction: generate $\{x_k^i, 1/N\}_{i=1}^N$ from $\{x_{k-1}^i, 1/N\}_{i=1}^N$ according to $p(x_k \mid x_{k-1})$

(2) Update: calculate the particle weights $\widetilde{w}_k^i = p(z_k \mid x_k^i)$ to obtain $\{x_k^i, \widetilde{w}_k^i\}_{i=1}^N$

(3) Normalize: calculate the total weight $t = \sum_{i=1}^N \widetilde{w}_k^i$ and normalize $w_k^i = \widetilde{w}_k^i / t$

(4) Resample: $\{x_k^i, w_k^i\}_{i=1}^N$ into a new uniformly weighted particle set $\{x_k^{i*}, 1/N\}_{i=1}^N$

Algorithm 1: SIR particle filter.

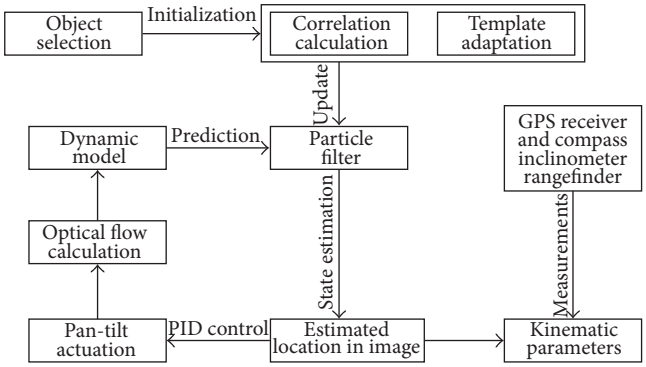


Figure 2: The schematic representation of the approach for assessing kinematic parameters of distant objects.

a well-established method for visual tracking or measuring displacements and deformations in image sequences [12, 13]:

$$M(x,y) = \frac{\sum_{x',y'} \left(T'(x',y') I'(x+x', y+y')\right)}{\sqrt{\sum_{x',y'} T'(x',y')^2 \sum_{x',y'} I'(x+x', y+y')^2}}, \tag{1}$$

where

$$T'(x',y') = T(x',y') - \frac{1}{w \cdot h} \sum_{x'',y''} T(x'',y''),$$

$$I'(x+x', y+y') = I(x+x', y+y') - \frac{1}{w \cdot h} \sum_{x'',y''} I(x+x'', y+y''). \tag{2}$$

In (1), M is the correlation metrics between the selected template T and the image region I at the location (x, y). In (2), w and h represent the width and height of the template T expressed in pixels.

However, the appearance of the object is expected to change over time due to changes in perspective, lighting conditions, occlusions, and so forth. To maintain the tracking ability, it is necessary to adapt the template according to the observed changes. We achieve this by evaluating the updated template [10] with a low-pass IIR filter with adaptive coefficients:

$$T_{k+1} = \begin{cases} \alpha N_k + (1-\alpha) T_k, & \text{if } M_{\max} > M_{\text{threshold}} \\ T_k, & \text{otherwise.} \end{cases} \tag{3}$$

In (3), T_k represents the current template, N_k represents the region in the current image with the highest correlation coefficient $M_{\max}$, T_{k+1} represents the updated template, and α represents the adaptive coefficient that is proportional to $M_{\max}$.

The adaptive template matching procedure yields the location of the region in current image that is the most similar to the current template. By repeating the procedure through the entire image sequence, we obtain a track. Since every obtained location is a measurement with a certain measurement uncertainty, the track can contain abrupt changes in location or velocity which are not necessarily consistent with the dynamic properties of the tracked object. To address this issue, we introduce a dynamic object model and use it in a particle filtering approach to recursively estimate the state of the tracked object as the new measurements become available.

We utilize sets of particles (or samples) to approximate probability density functions that are used in Bayesian filtering as estimates of the system state (at time k) x_k which is the location of the tracked object in the image and is represented by a set of weighted particles $\{x_k^i, w_k^i\}_{i=1}^N$ where w_k^i is the weight of the ith particle x_k^i. The task of the particle filter is to recursively estimate the position of the tracked object based on the measurements (template matching) and predictions of the dynamic model. A single iteration of the filtering procedure is carried out in two steps, namely, prediction and update. The prediction uses all the information available up to the time $k-1$ and the dynamic model to predict the system state at time k. When a new measurement becomes available, this information is used to update the system state. The employed particle filtering approach is presented in Algorithm 1.

The probability distribution $p(x_k \mid x_{k-1})$ presents the transitional prior which is defined by the dynamic model. We utilize a constant velocity autoregressive dynamic model to account for the inertia of the moving vehicles, while we model the movement of people as Brownian motion. However, in

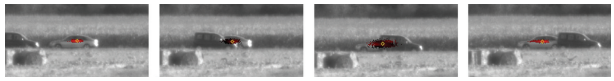


Figure 3: Vehicle tracking during occlusion.

certain cases, the constant velocity model is also suitable for tracking people [5, 14].

The probability distribution $p(z_k \mid x_k)$ utilized in the update step presents the likelihood function which is evaluated for each particle location using the presented correlation metrics between the template and the image region.

In the implementation of particle filtering, we use a version of SIR (sequential importance resampling) filter [15] which introduces the resampling step at every iteration to avoid the degeneration of the particle set. By resampling, the particles with small weights are removed, while the ones with high weights are multiplied to obtain an equally weighted particle set. In this way, the resampling step moves the particles towards the regions of high likelihood which are determined with the template matching procedure. Nevertheless, in the case of occlusion, the measurements based on template matching are irrelevant. For that reason, the resampling is not performed and the particles move according to the dynamic model, which increases the probability of restoring the track once the occlusion has passed. This is demonstrated in Figure 3 by an example of tracking a vehicle at a distance of about 1 km.

In Figure 3, the tracking is performed with a set of 1000 particles that are represented as colored points. The color of each particle depends on its weight that is determined by the template matching procedure in the update step of the particle filter. The yellow circle represents the estimated location of the tracked vehicle and is calculated as the weighted average of the particle set.

During the process of tracking, the camera and the laser rangefinder move in order to maintain the orientation towards the moving object. To achieve this, we utilize a PID control algorithm where the estimated location of the distant object in the image is used to control the actuation of the pan-tilt unit. This movement introduces a substantial amount of vibration that decreases the image quality and limits the tracking ability of the system. In order to overcome this issue, we exploit the high sensitivity of the employed camera and reduce the exposure time to 1 ms. Another difficulty with the moving camera is that we need to properly characterize its movement in order to determine the kinematic parameters of the tracked object in a global coordinate system. This is achieved by the calculation of the optical flow of the background with an implementation of the Lucas-Kanade method [11].

3. Experiments

The experimental verification of the developed method and system was conducted in an open field to ensure adequate visibility of the tracked objects and proper operation of the GPS receiver. The assessment of the kinematic parameters was performed for moving people at distances of 500 m to 1 km and moving vehicles at distances of 1 km.

To control the measuring process we employed a personal computer with Intel Core 2 CPU and 1 GB memory. During the experiments, the camera was acquiring 17 frames per second at the resolution of 672 × 448 pixels while the laser rangefinder was performing distance measurement with the period of 6 s. The frame rate of the image acquisition was limited due to the recording of the acquired images to the hard disk of the system computer for the means of documentation. Without the recording, the achievable frame rate (limited by the camera) was 25 frames per second. The employed image processing algorithms alone (with the use of 1000 particles) are able to process 50 images per second.

3.1. Measurements of Vehicle Kinematics. To characterize the movement of a vehicle, we selected a road section that is positioned in lateral direction with respect to the position of the measuring system. This enables us to assess the performance of the visual measuring system and partially eliminate the effects of low measurement resolution and repetition rate of the employed laser rangefinder.

In the presented experiment the vehicle accelerated from 0 to 50 km/h, then started decelerating and finally disappeared behind obstacles. During the experiment, the vehicle was tracked from the distance of approximately 1 km by utilizing the presented measurement system. Since the outcomes of the visual tracking are expressed in pixels, we convert them to angles using the field of view per pixel which is calculated from the pixel size and the focal length of the camera lens.

The results of the optical flow measurements are displayed in Figure 4(a) as horizontal and vertical components of the camera angular velocities. By employing these measurements and the location of the tracked vehicle in the image, estimated by the particle filter, we obtain the angular velocities of the vehicle (Figure 4(b)).

We notice that the vibrations of the camera manifested as fluctuations of angular velocity in Figure 4(a) are not transferred to the measured angular velocities of the vehicle (in Figure 4(b)). This means that the optical flow algorithm successfully compensates the mechanical vibrations that are produced by the actuators in the pan-tilt unit.

During the experiment, the distance to the moving vehicle was repeatedly measured by the laser rangefinder. Knowing of the distance enables the determination of the vehicle velocity with respect to the surroundings which is presented in Figure 5. To enable the evaluation of the characteristics of the remote measurement system we equipped the tracked vehicle with a high performance GPS receiver and a portable computer for recording the positions and velocities. The comparison of velocities acquired by the two set-ups shows good agreement. The calculated RMSE (root mean square error) before the total occlusion is 0.9 km/h. The time resolution of the velocity measurements, which is limited by the frame rate of the employed camera, is high enough to allow detection of shifting to higher gears during acceleration.

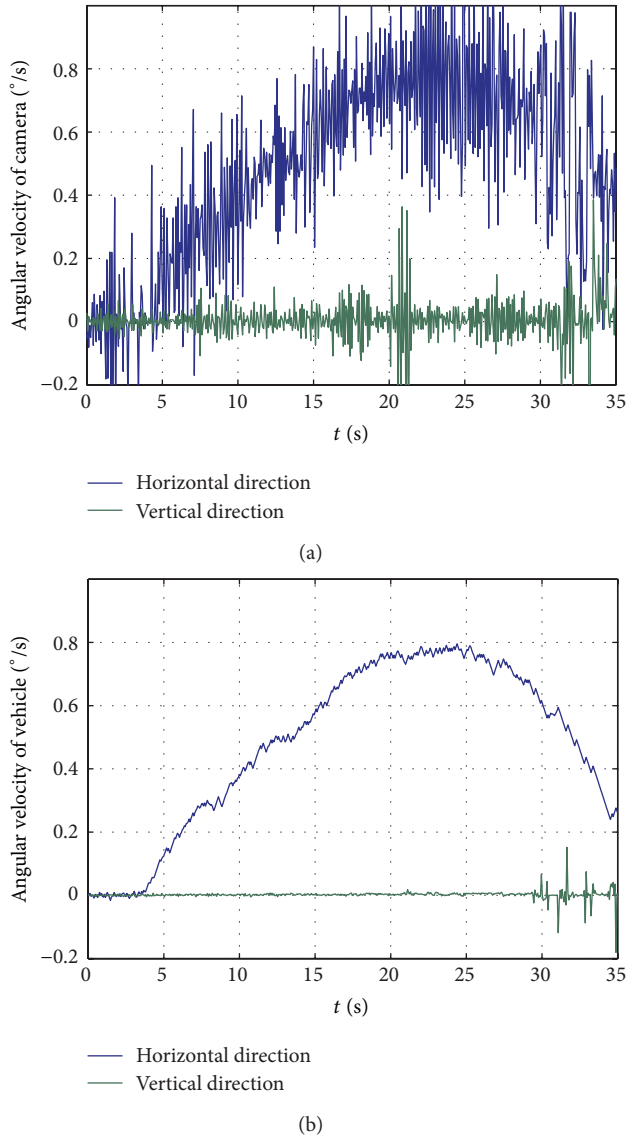


Figure 4: (a) Angular velocities of the camera and (b) the tracked vehicle.

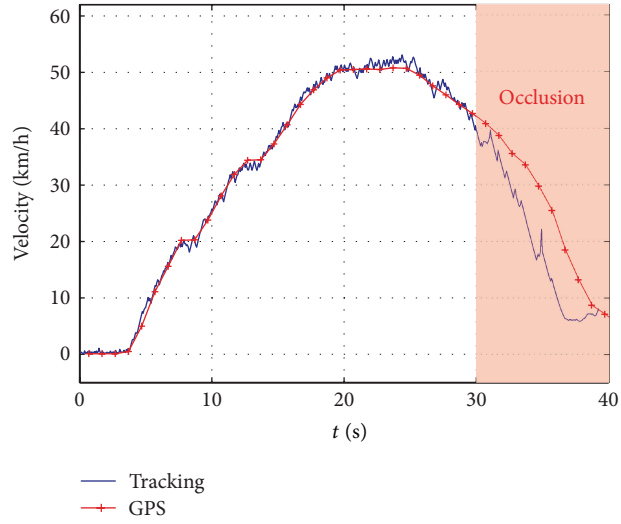


Figure 5: Velocity of the tracked vehicle.

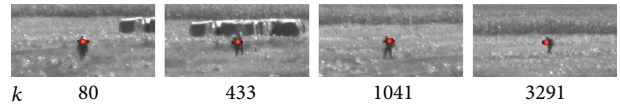


Figure 6: Selected frames acquired during tracking a moving person.

3.2. Measurements of People Kinematics. To investigate the possibilities for characterizing the people movement with the developed measuring system, we performed several experiments with a person walking in different directions at distances above 500 m.

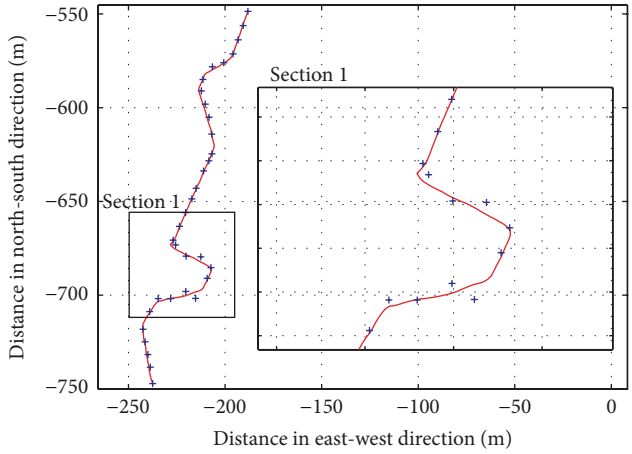
In Figure 6, we show a person that walks in various directions and slowly moves away from the measuring system. The parameter k below the frames represents the sequential frame number. We notice that the tracking is successful until the frame 1041. In later frames, the particles begin to move away from the tracked person, which is evident in frame 3291. This is caused by the template drift which is common in adaptive template matching and has been the topic of many studies [16, 17].

The obtained results clearly demonstrate the need to investigate the options for prevention of the template drift since it limits the ability to perform laser distance measurements after longer times. If the set of particles moves from the person, the laser rangefinder is unable to measure the distance to the person and measures the distance to the background instead.

With the present set-up we are able to track a moving person for several minutes before the template drift becomes too large to perform further measurements. An example of such tracking session is presented in Figure 7. Instead of the measured velocities, we show the obtained trajectory in horizontal plane (Figure 7(a)) and on a 3D map in Google Earth (Figure 7(b)). In the experiment, the tracked person carried a backpack with a high performance GPS receiver and a portable computer to record the positions and velocities. The GPS trajectory is also shown in Figures 7(a) and 7(b) to enable the evaluation of the performance of the developed measuring system.

The comparison of the trajectories in Figure 7(a) shows overall good agreement with few characteristic discrepancies, which become notable when the person moves in a tangential direction as can be seen in the enlarged section of the diagram. These differences are caused primarily by the distance measurement uncertainty which is about ±3 m (95% confidence level). Nevertheless, the calculated RMSE with respect to the reference GPS measurements is only 2.0 m.

Another set of experiments was performed at distances of approximately 1 km. The results showed that the tracking was successful only for a period of about 80 seconds. This is attributed to the effect of the template drift which causes the estimated location of the object in the image to move away



(a)

(b)

FIGURE 7: Trajectory of a moving person measured by means of the tracking system (symbols) and a GPS receiver (solid line) displayed in the horizontal plane (a) and on a 3D map (b).

from the actual location. The template drift eventually results in improper distance measurements which deteriorate the tracking performance. This issue is especially evident when tracking small objects at long distances.

4. Conclusion

The presented results demonstrate the possibility for characterization of the movement of various noncooperative objects at the distances up to 1 km, by acquiring their velocity and trajectory in real time. The developed experimental system integrates a laser rangefinder, a high performance CCD camera and appropriate image processing techniques to enable measurements over long distances in diverse terrains. The specific implementation determines the trajectory in a global coordinate system (WGS 84), allowing a straightforward display in a geographic information system. The key limitations of the current experimental set-up are low repetition rate and resolution of range measurements and template drift that occurs in longer tracking sessions.

References

[1] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 34, no. 3, pp. 334–352, 2004.

[2] H. Zhao and R. Shibasaki, "A novel system for tracking pedestrians using multiple single-row laser-range scanners," *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 35, no. 2, pp. 283–291, 2005.

[3] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809–830, 2000.

[4] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking groups of people," *Computer Vision and Image Understanding*, vol. 80, no. 1, pp. 42–56, 2000.

[5] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Tracking multiple people using laser and vision," in *Proceedings of the IEEE IRS/RSJ International Conference on Intelligent Robots and Systems (IROS '05)*, pp. 1301–1306, August 2005.

[6] Y. Goyat, T. Chateau, and L. Trassoudaine, "Tracking of vehicle trajectory by combining a camera and a laser rangefinder," *Machine Vision and Applications*, vol. 21, no. 3, pp. 275–286, 2010.

[7] A. Fod, A. Howard, and M. J. Matarić, "A laser-based people tracker," in *Proceedings of the IEEE International Conference on Robotics adn Automation (ICRA '02)*, vol. 3, pp. 3024–3029, May 2002.

[8] L. Kuscer and J. Diaci, "Measurement uncertainty assessment in remote object geolocation," *Journal of Mechanical Engineering*, vol. 59, no. 1, pp. 32–40, 2013.

[9] R. Burtch, "A comparison of methods used in rectangular to geodetic coordinate transformations," in *Proceedings of the American Congress on Surveying and Mapping*, pp. 12–14, 2006.

[10] J. Ahmed, M. N. Jafri, M. Shah, and M. Akbar, "Real-time edge-enhanced dynamic correlation and predictive open-loop car-following control for robust tracking," *Machine Vision and Applications*, vol. 19, no. 1, pp. 1–25, 2008.

[11] G. Bradski and A. Kaehler, *Learning OpenCV*, O'Reilly Media, Sebastopol, Calif, USA, 2008.

[12] B. Pan, K. Qian, H. Xie, and A. Asundi, "Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review," *Measurement Science and Technology*, vol. 20, no. 6, Article ID 062001, pp. 1–17, 2009.

[13] D. Das, C. L. Dora, D. Saravanan, and K. Karunakar, "Characteristics of embedded-shock-free compressible vortex rings: a detailed study using PIV," *Advances in Mechanical Engineering*, vol. 2011, Article ID 650871, 13 pages, 2011.

[14] K. O. Arras, S. Grzonka, M. Luber, and W. Burgard, "Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '08)*, pp. 1710–1715, May 2008.

[15] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-gaussian Bayesian state estimation," *IEE Proceedings F*, vol. 140, no. 2, pp. 107–113, 1993.

[16] P. Jiyan and H. Bo, "Robust object tracking against template drift," in *Proceedings of the 14th IEEE International Conference on Image Processing (ICIP '07)*, vol. 3, pp. III353–III356, September 2007.

[17] B. Fan, Y. Du, Y. Cong, and Y. Tang, "Active drift correction template tracking algorithm," in *Proceedings of the19th IEEE International Conference on Image Processing (ICIP '12)*, pp. 397–400, 2012.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 976864, 8 pages
http://dx.doi.org/10.1155/2013/976864

Research Article

Development of Lifting and Propulsion Mechanism for Biped Robot Inspired by Basilisk Lizards

Linsen Xu,[1,2] Tao Mei,[1] Xianming Wei,[3] Kai Cao,[2] and Mingzhou Luo[1]

[1] *Hefei Institute of Physical Science, Chinese Academy of Science, Hefei 230031, China*
[2] *Changzhou Institute of Advanced Manufacturing Technology, Changzhou 213164, China*
[3] *University of Science and Technology of China, Hefei 230026, China*

Correspondence should be addressed to Linsen Xu; xls008@126.com

Received 19 March 2013; Accepted 21 September 2013

Academic Editor: Hongxing Wei

The lifting and propulsion mechanism of a novel biped robot inspired by the basilisk lizard's water-walking function has been developed. The movement trajectories of the Watt-I planar linkage are brought out by combining the movement equations of the four-bar mechanism and the coordinate transformation equations, which are used to simulate the foot trajectories of the basilisk lizard, and the lifting and propulsion mechanism of the biped robot walking on water is carried out. The links' parameters are optimized by taking the trajectories overlap ratio as the objective function. The prototype of the biped robot walking on water is manufactured by the results of the kinematic analysis on the robot. And the lifting and propulsion force curve on the robot from water is measured. The experiment results show that the lifting and propulsion system can satisfy the function requirement of the biped robot walking on water.

1. Introduction

The requirement that the robot has the functions of walking on both land and water is brought out to implement the works such as military surveillance, water quality monitoring, wetland detection, search, and rescue in the complex environments. The technologies of the legged robot walking on the hard land surface are mature relatively [1, 2], so it has been the research hotspot in the robotics that study the mechanisms of legged robots walking on different kinds of material surface such as water and soft sand to expand its walk domain.

Song and Sitti have developed a robot with twelve legs which can walk on water by mimicking the water skipper [3]. The robot floats by the water surface tension and moves by driving the special paddle-type legs, so the payload of the robot is low (9.3 g), and its walk velocity is slow (3 cm/s). The robot will sink when the water-air interface which keeps its balance broken, which means that the robot cannot walk on the water surface with violent ups and downs, and its ability of resisting windstorm is weak.

Floyd and Sitti has developed a quadruped robot walking on water, which realizes the water-walking function by its four disk feet striking water [4]. The robot feet move with the constant angle, which leads to the bigger water resistance when they are lifted and reduces the robot's payload (50 g).

The basilisk lizard is capable of walking on water surface at the speed of about 1.5 m/s and at a stepping rate of 5–10 Hz (per leg). Compared with a ship, the lizard's move mode leads to the smaller volume under water and can reduce the water resistance, which can promote the propeller efficiency. And the biped water-walking mode has the following advantages.

(1) The water resistance of the biped walking mode is smaller than the quadruped mode.

(2) Just like the human walks on the soft sand surface, the biped water-walking mode can make the lizard body lean forward, which is useful to control the body's balance.

(3) The biped water-walking mode can make the lizard head be higher and get a more wilder field of vision.

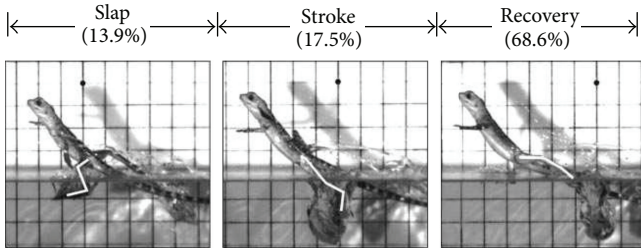


Figure 1: Pictures of basilisk lizard walking on water.

In this case, the water-walking ability is what our robot attempts to duplicate. This robot employs momentum transfer for both lifting and propulsion, instead of surface tension, which other water-walking robots adopt.

The work is helpful to expand the locomotion domain of biped robot. Further work in this field can lead to completely amphibious bipedal motion. Applications include exploration, and search, and rescue in partially flooded or marsh-like environments and of remote controlled toy models which can run anywhere. This work can also help increase the understanding of the basilisk lizard and its ability to walk on both land and water.

In this paper, we develop a novel biped water-walking robot inspired by basilisk lizards. Firstly, we emulate the water-walking function of the basilisk lizard by Watt-I planar linkage. Then the lifting and propulsion mechanism is developed, and the kinematics analysis of the mechanism is carried out to select the suitable body material and driver. To obtain the lizard's joints trajectories as more exact as possible, we optimize the links' parameters. At last we manufacture the biped robot prototype to test its lifting and propulsion ability.

2. Research on Dynamics Mechanism of Biped Robot Walking on Water

The biped robot walking on water is aim to the water-running function of basilisk lizards. The water-walking stride can be roughly divided into three phases: slap, stroke, and recovery. The forces experienced by the leg and foot are different in each phase and have differing effects on the lizard's ability to stay afloat. These phases are shown in Figure 1. The surface-tension effect on the water-walking ability is negligible. It can be known from Figure 2 that the joints trajectories on one leg of a lizard are in the same plane approximately.

In this paper, we implement the water-walking function of the biped robot by simulating the sole trajectory of the lizard walking on water, so the robot's water-walking stride is also divided into three phases: slap, stroke, and recovery. In the slap and stroke phase, the lizard sole strokes downward, pushes against the water beneath its foot, and create an air cavity in the water. The momentum transfer from the lizard's foot to the water during this stroke phase generates the main lift force to stay afloat and the main propulsion force to thrust forward. From [5, 6], the force of lizard foot entering water is

$$F_{\text{str}}(t) = C_D^* \left[0.5S\rho \cdot v^2 + S\rho \cdot gh(t)\right], \quad (1)$$

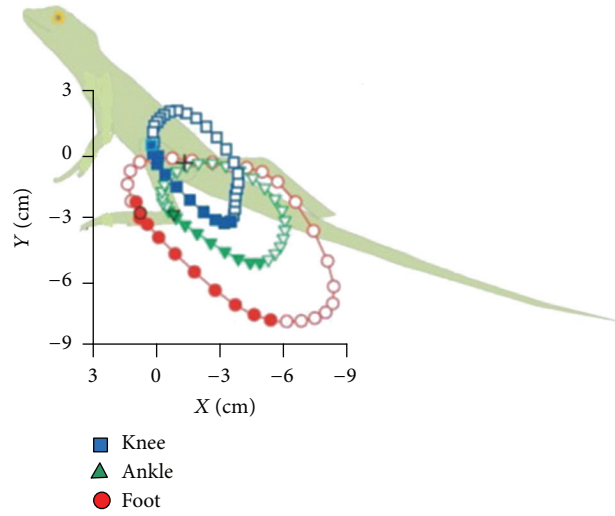


Figure 2: Movement trajectories of basilisk lizard walking on water.

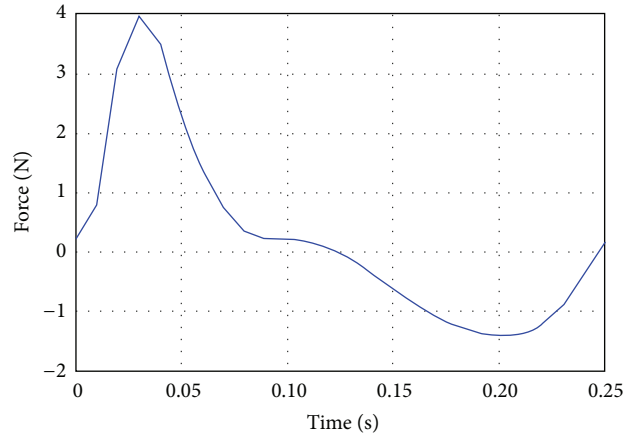


Figure 3: Force curve during robot walking on water in simulation environment.

where $C_D^* \approx 0.703$ is the constant drag coefficient, ρ is the water density, g is the acceleration of gravity, S is the effective area, and $h(t)$ is the time varying depth of the foot. This holds true over a large range of velocities for both lizards and experimental equipment. The force curve in a stride can be got according to the above equation, which is shown in Figure 3.

3. Design of Lifting and Propulsion Mechanism

3.1. Kinematics Analysis of Double Bar Assur Group of Watt-I Linkage. Watt-I linkage is shown in Figure 4, and the double bar Assur Group links the connecting link and the output link of the grounded four-bar linkage. The double bar Assur

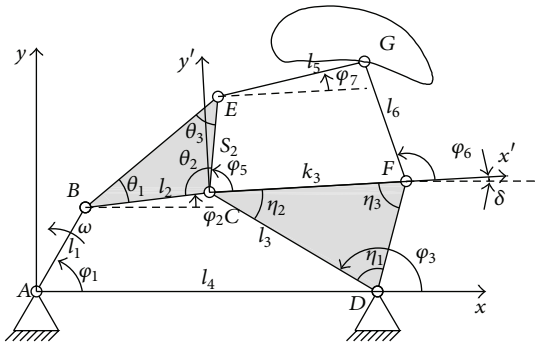


Figure 4: Watt-I planar linkage.

Group will not constrain the four-bar linkage as long as the following conditions are satisfied:

$$\begin{aligned} \max|EF| &\le |EG| + |FG|, \\ ||EG| - |FG|| &\le \min|EF|. \end{aligned} \tag{2}$$

In Figure 4, the connecting links of AB, BC, CD, and AD constitute the grounded four-bar linkage, whose lengths are represented by l_1, l_2, l_3, and l_4, respectively. Suppose that the angles between the links AB, BC, and CD and the x-axis are φ_1, φ_2, and φ_3, respectively, and the interior angles of ΔBCE and ΔDCF are θ_i and φ_i $(i = 1, 2, 3)$, respectively. The link AB is the driving link at the constant angular velocity of ω. So the angular displacements can be known as follows respectively [7]:

$$\begin{aligned} \varphi_3 &= 2\tan^{-1}\frac{N + \sqrt{M^2 + N^2 - Q^2}}{M - Q}, \\ \varphi_2 &= \tan^{-1}\frac{N + l_3\sin\varphi_3}{M + l_3\cos\varphi_3}, \end{aligned} \tag{3}$$

where $M = l_4 - l_1\cos\varphi_1$, $N = -l_1\sin\varphi_1$, $Q = (M^2 + N^2 + l_3^2 - l_2^2)/2l_3$, and $\varphi_1 = \omega t$.

So the trajectory of the point C can be got from (3):

$$\begin{pmatrix} x_C \\ y_C \end{pmatrix} = \begin{pmatrix} l_1\cos\varphi_1 + l_2\cos\varphi_2 \\ l_1\sin\varphi_1 + l_2\sin\varphi_2 \end{pmatrix}. \tag{4}$$

The mechanism of $ABCD$ is a quadrilateral, so

$$\begin{aligned} \angle BCD &= 2\pi - \varphi_1 - (\varphi_2 + \pi - \varphi_1) - (\pi - \varphi_3) \\ &= \varphi_3 - \varphi_2. \end{aligned} \tag{5}$$

Supposing $\angle ECF = \varphi_5$, its value can be deduced out from (5):

$$\varphi_5 = 2\pi - \theta_2 - \eta_2 - (\varphi_3 - \varphi_2). \tag{6}$$

A new coordinate system $x'Cy'$ is built by setting the point C as the origin and the side CF as the x'-axis, so the angle between the x'-axis and the x-axis is got as

$$\begin{aligned} \delta &= \pi - \eta_3 - (\pi - (\varphi_3 - \eta_1)) \\ &= \varphi_3 - \eta_3 - \eta_1. \end{aligned} \tag{7}$$

The links of CE, EG, GF, and CF constitute a four-bar linkage also, and the lengths are represented as s_2, l_5, l_6, and k_3, respectively. The link CE is looked as the driving link. Supposing that the angles between the rods EG and FG and the x'-axis are φ_7 and φ_6, respectively, the angular displacements can be known as

$$\begin{aligned} \varphi_6 &= 2\tan^{-1}\frac{J + \sqrt{H^2 + J^2 - K^2}}{H - K}, \\ \varphi_7 &= \tan^{-1}\frac{J + l_6\sin\varphi_6}{H + l_6\cos\varphi_6}, \end{aligned} \tag{8}$$

where $H = k_3 - s_2\cos\varphi_5$, $J = -s_2\sin\varphi_5$, and $K = (H^2 + J^2 + l_6^2 - l_5^2)/2l_6$.

So the trajectory of the point G in the coordinate system $x'Cy'$ can be deduced out:

$$\begin{pmatrix} x'_G \\ y'_G \end{pmatrix} = \begin{pmatrix} s_2\cos\varphi_5 + l_5\cos\varphi_7 \\ s_2\sin\varphi_5 + l_5\sin\varphi_7 \end{pmatrix}. \tag{9}$$

And the trajectory of the point G in the coordinate system xAy can be got according to the transformation formulas of the coordinates:

$$\begin{pmatrix} x_G \\ y_G \end{pmatrix} = \begin{pmatrix} x'_G\cos\delta - y'_G\sin\delta + x_C \\ y'_G\sin\delta + y'_G\cos\delta + y_C \end{pmatrix}. \tag{10}$$

The trajectory of the point G is shown in Figure 4.

3.2. Virtual Prototype Design of Lifting and Propulsion Mechanism. The trajectory of the point G shown in Figure 4 is similar with the sole trajectory of the basilisk lizard shown in Figure 2, so we adopt the improved Watt-I mechanism as the lifting and propulsion mechanism for the biped robot.

We look on the point G as the center point of the robot's sole and move it to be under the x-axis to implement the stroke function of the lifting and propulsion mechanism. Simultaneously the triangle plates are replaced by the slender rods, so $\theta_1 = \pi$, $\eta_1 = 0$, $\theta_2 = \theta_3 = \eta_2 = 0$, $\eta_3 = \pi$, and

$$\begin{aligned} \varphi_5 &= 2\pi - \varphi_3 + \varphi_2, \\ \delta &= \varphi_3 - \pi. \end{aligned} \tag{11}$$

Moreover, we rotate the rod AD an angle less than $\pi/4$ around point A to make the sole parallel to the water surface, then the lifting and propulsion force may be bigger. The final skeleton of the lifting and propulsion mechanism is shown in Figure 5.

It can be known from Figure 1 that the two feet slap the water surface alternately during the lizard walking on water. For mimicking the lizard walking on water, we design the same lifting and propulsion mechanisms for the both sides of the biped robot. The mechanisms are driven by a same motor, but their phase difference is π. The major structure of the biped robot is shown in Figure 6.

In Figure 6, the driving system is composed of a motor and a reducer fixed on the main frame, and the transmission

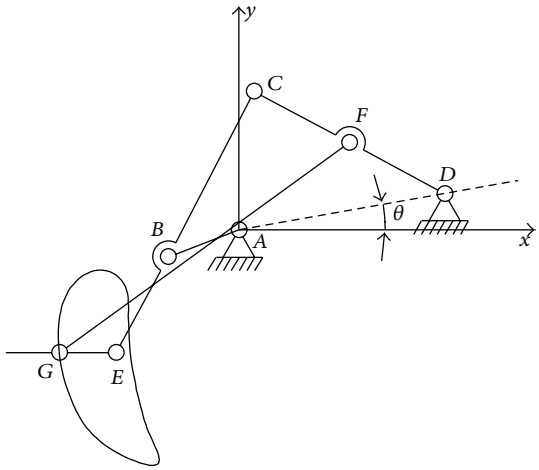


Figure 5: Skeleton of lifting and propulsion mechanism.

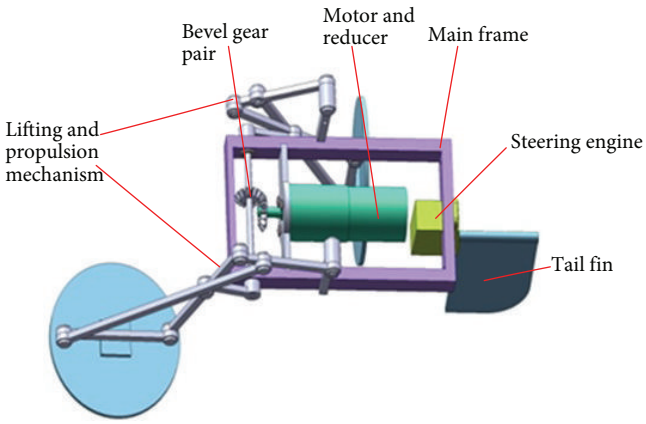


Figure 6: Major structure of robot.

agent includes a bevel gear pair and a transmission shaft. The bevel gear pair turns the driving motion perpendicular to the rotational motion around the shaft, and the driving torque is passed to the two lifting and propulsion mechanisms. A balance apparatus is fitted on the after body of the robot's frame, which can keep the robot being balance when it is walking on water.

3.3. Optimization of Lifting and Propulsion Mechanism. To imitate the water-walking function of the lizard perfectly, the trajectory of point G should fit the points on the given trajectory as more as possible. The given trajectory parameters during the lizard slap and stroke phases are shown in Table 1.

The optimized objective function for reappearing the trajectory is built as follows according to minimizing the deviation between the points of the actual trajectory and the given:

$$f(x) = \sum_{i=1}^{n} w_i \left[\left(O_{xi} - \dot{O}_{xi} \right)^2 + \left(O_{yi} - \dot{O}_{yi} \right)^2 \right] \longrightarrow \min, \tag{12}$$

Table 1: Parameters of given trajectory. Unit: mm.

Number	1	2	3	4	5	6	7
x_i	−66.5	−65.6	−64.1	−61.6	−59.2	−56	−52.7
y_i	−34.9	−37.9	−41.1	−44.5	−48.1	−51.9	−55.6
No	8	9	10	11	12	13	14
x_i	−49.4	−45.2	−40.4	−35.3	−30.4	−26	−21
y_i	−59.4	−63	−66.5	−69.7	−72.5	−74.9	−76.7

where $\dot{O}_{xi}$, $\dot{O}_{yi}$ $(i = 1, 2, \ldots, 14)$ are the coordinates of the points on the given trajectory, O_{xi}, O_{yi} $(i = 1, 2, \ldots, 14)$ are the coordinates of the points on the actual trajectory, and w_i $(i = 1, 2, \ldots, 14)$ are the preselected weighted factors which represent the importance degree of each point deviation.

There are two constraint conditions for the four-bar linkage $ABCD$ shown in Figure 4: (1) the constraint condition of crank existence in a four-bar linkage and (2) the constraint condition of the transmission angle $[\zeta]$. We set $[\zeta_{\min}] \geq 25^\circ$ and $[\zeta_{\max}] = 180^\circ - [\zeta_{\min}]$, so the constraint conditions can be written as

$$\begin{aligned} g_1(x) &= AB > 0, \\ g_2(x) &= AD + CD - BC - AB \geq 0, \\ g_3(x) &= AD + BC - CD - AB \geq 0, \\ g_4(x) &= BC + CD - AD - AB \geq 0, \\ g_5(x) &= \cos\zeta_{\max} - \cos\left|\zeta_{\max}\right| \geq 0, \\ g_6(x) &= \cos\zeta_{\min} - \cos\left|\zeta_{\min}\right| \geq 0, \end{aligned} \tag{13}$$

where

$$\begin{aligned} \cos\zeta_{\max} &= \frac{BC^2 + CD^2 - (AB + AD)^2}{2BC \cdot CD}, \\ \cos\zeta_{\min} &= \frac{BC^2 + CD^2 - (AB - AD)^2}{2BC \cdot CD}. \end{aligned} \tag{14}$$

Moreover, it can be known from Figure 2 that the double bar Assur Group should satisfy the following conditions to keep the mechanism working steadily.

(1) When the stability margin is 20°, the condition of the mechanism working without a singular point should be

$$\max\left(\varphi_7 + \delta - \varphi_2\right) \leq 160^\circ. \tag{15}$$

(2) The condition of the sole slapping the water surface effectively to produce the lifting and the propulsion force should be

$$\varphi_7 + \delta > 90^\circ. \tag{16}$$

For minimizing the deviation between the points of the actual trajectory and the given, we set (12) as the objective

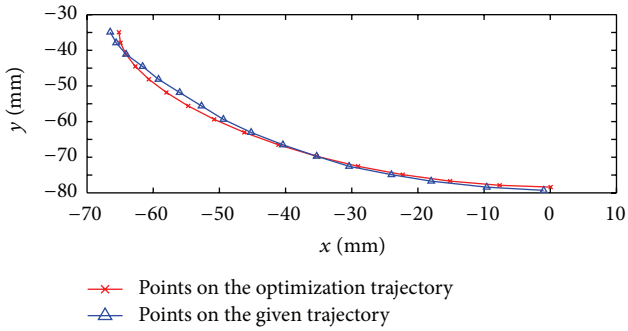


FIGURE 7: Comparison between parameters of optimization trajectory and given trajectory.

function, (3), (8), and (10) as the variable functions, and (13)~(16) as the constraint conditions. The optimum values of the design variables can be got as follows by taking the penalty function method: $AB = 24.53$ mm, $CF = 12.69$ mm, $DF = 35.58$ mm, $BC = 38.64$ mm, $BE = 39.31$ mm, $FG = 114.64$ mm, $EG = 54.06$ mm, $AD = 49.76$ mm, and $\delta = 42.83^\circ$.

The comparison between the parameters of the optimization trajectory and the given trajectory is shown in Figure 7.

4. Kinematics Analysis on Robot

The velocities of the links CD and BD can be deduced out based on the kinematics principle of a four-bar mechanism, which are as follows:

$$\omega_3 = \omega \frac{l_1 \sin(\varphi_1 - \varphi_2)}{l_3 \sin(\varphi_3 - \varphi_2)},$$
$$\omega_2 = -\omega \frac{l_1 \sin(\varphi_1 - \varphi_3)}{l_2 \sin(\varphi_2 - \varphi_3)}. \tag{17}$$

And the angular accelerations of the links CD and BC are, respectively,

$$\alpha_3 = \frac{l_2\omega_2^2 + l_1\omega_1^2 \cos(\varphi_1 - \varphi_2) - l_3\omega_3^2 \cos(\varphi_3 - \varphi_2)}{l_3 \sin(\varphi_3 - \varphi_2)},$$
$$\alpha_2 = \frac{l_3\omega_3^2 - l_1\omega_1^2 \cos(\varphi_1 - \varphi_3) - l_2\omega_2^2 \cos(\varphi_2 - \varphi_3)}{l_2 \sin(\varphi_2 - \varphi_3)}. \tag{18}$$

The angular velocity and acceleration of φ_5 can be worked out from (11), and they are, respectively,

$$\omega_5 = -\omega_3 + \omega_2,$$
$$\alpha_5 = -\alpha_3 + \alpha_2. \tag{19}$$

In the same way, the angular velocity and acceleration of δ are, respectively,

$$\omega_\delta = \omega_3,$$
$$\alpha_\delta = \alpha_3. \tag{20}$$

According to the kinematics principle of a four-bar mechanism, the angular velocities of the links GF and EG in the coordinate system $x'Cy'$ are known as

$$\omega_6' = \omega_5 \frac{s_2 \sin(\varphi_5 - \varphi_7)}{l_6 \sin(\varphi_6 - \varphi_7)},$$
$$\omega_7' = -\omega_5 \frac{s_2 \sin(\varphi_5 - \varphi_6)}{l_5 \sin(\varphi_7 - \varphi_6)}, \tag{21}$$

and the values of their angular acceleration in the coordinate system $x'Cy'$ are

$$\alpha_6' = \frac{l_5\omega_7'^2 + s_2\omega_5'^2 \cos(\varphi_5 - \varphi_7) - l_6\omega_6'^2 \cos(\varphi_6 - \varphi_7)}{l_6 \sin(\varphi_6 - \varphi_7)},$$
$$\alpha_7' = \frac{l_6\omega_6'^2 - s_2\omega_5'^2 \cos(\varphi_5 - \varphi_6) - l_5\omega_7'^2 \cos(\varphi_7 - \varphi_6)}{l_5 \sin(\varphi_7 - \varphi_6)}. \tag{22}$$

Then the angular velocities and accelerations of the links GF and EG can be got from (20)~(22):

$$\omega_6 = \omega_3 + \omega_6'$$
$$\omega_7 = \omega_3 + \omega_7',$$
$$\alpha_6 = \alpha_3 + \alpha_6',$$
$$\alpha_7 = \alpha_3 + \alpha_7'. \tag{23}$$

5. Manufacture of Prototype and Experiments

5.1. Design of Control System. The ZMP criterion is adopted by most of the biped robot walking on land to keep balance. But when a robot is walking on water, the air caves will come into being, and the disturbances of current rush and splash are more complex than on land. So a new control method is necessary to the biped robot walking on water.

It is necessary to reduce the foot slapping frequency properly to enhance the robot stability when it is disturbed. Therefore, the CPG-fuzzy control algorithm of the biped robot walking on water is proposed for achieving sufficient balance control and gait switching [8]. The CPG-fuzzy control system (Figure 8) of the robot is designed by using the CPG controller and the fuzzy control rules.

5.2. Manufacture of Prototype. In consideration of the functional requirements of the robot walking on water, the main frame, the lifting, and propulsion mechanisms should be made from the light material, and their intensity should be high enough to deform slightly under the action of the reaction force. So they should to be made from aluminum alloy LY12. The driver group MAXON 221011+134161 is adopted as the driving system of the biped robot. The rated power of the group is 5 w, and its rated torque is 0.15 Nm. We choose flexible rubber as the material of the robot soles. The prototype of the biped robot walking on water is shown in Figure 9, and its whole weight is 3.2 N.

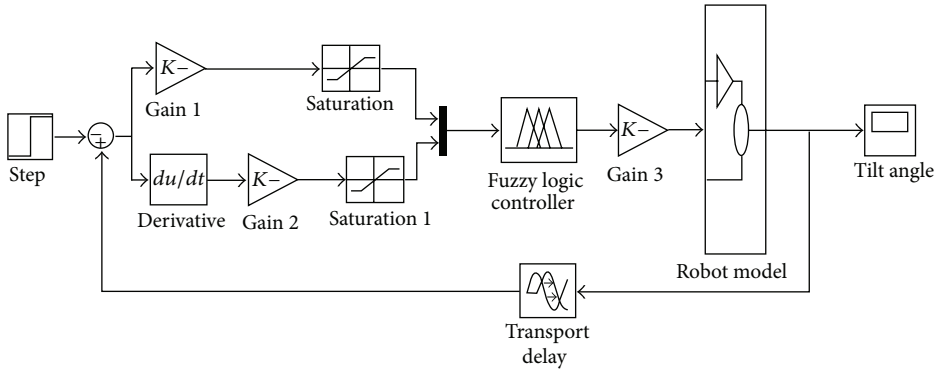


Figure 8: CPG-fuzzy control system of biped robot.

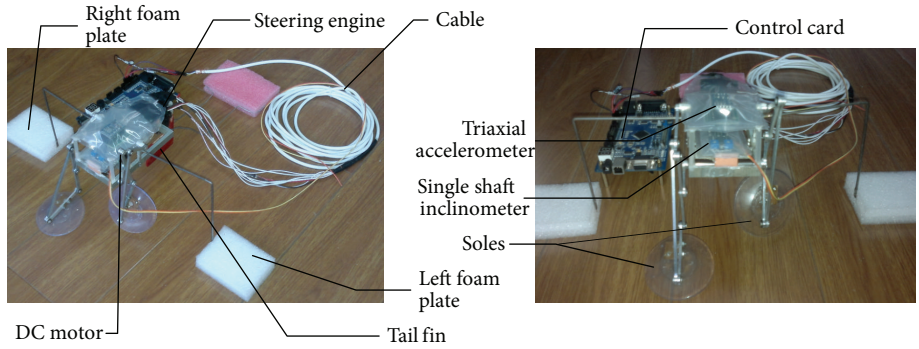


Figure 9: Prototype of biped robot walking on water.

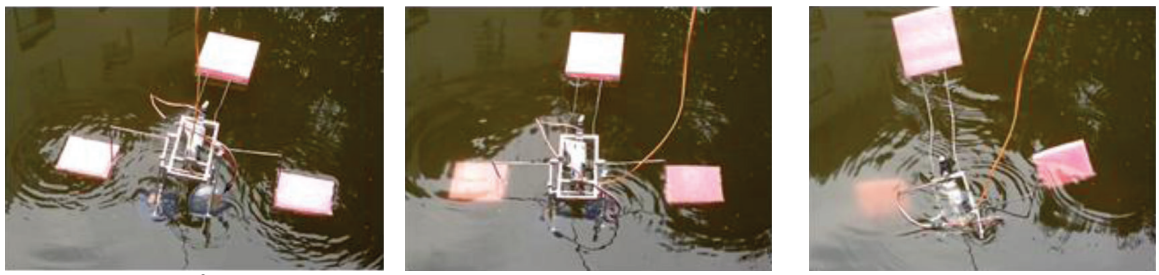


Figure 10: Sequence diagrams of robot out of work sinking.

5.3. Experiments. The volumes of the left foam plate and the right one are 91.605 cm^3 and 97.712 cm^3, respectively, and the total buoyancy of the two foam plates is 1.856 N when they are all under water. The buoyancy is less than its gravity, so the robot will sink when it is placed on the water statically, and the sequence diagrams of the robot sinking are shown in Figure 10.

The sequence diagrams of the robot walking on water are shown in Figure 11. The robot can walk on water due to the lifting and propulsion force produced by the legs slapping from 1 s to 4 s. As soon as the power supply is cut off when the time is 4 s, the robot will stop slapping and begin to sink in water. At last the robot will sink totally when the time is 6 s.

It can be known from Figures 10 and 11 that the lifting and propulsion force is the main factor to keep the robot away from sinking.

5.4. Measurement of Lifting and Propulsion Force. The lifting and propulsion force is the decisive factor of the robot's load capacity; that is, if the force is larger, its load capacity will be stronger. We build an experimental platform shown in Figure 12 to test the lifting and propulsion force of the robot. We mount the strain gages on both soles (Figure 13), and the voltage variation curve of one gage is shown by an oscilloscope, from which the lifting and propulsion force can be got. And the curve of the lifting and propulsion force is shown in Figure 14.

The force curve of Figure 14 is similar to the curve in the simulation environment shown in Figure 3.

The peak value of the lifting and propulsion force in Figure 14 is 2.6 N, which is less than the force value of 4 N in the simulation environment. And the valley value of the force in Figure 14 is about 0.8 N, which is less than the value

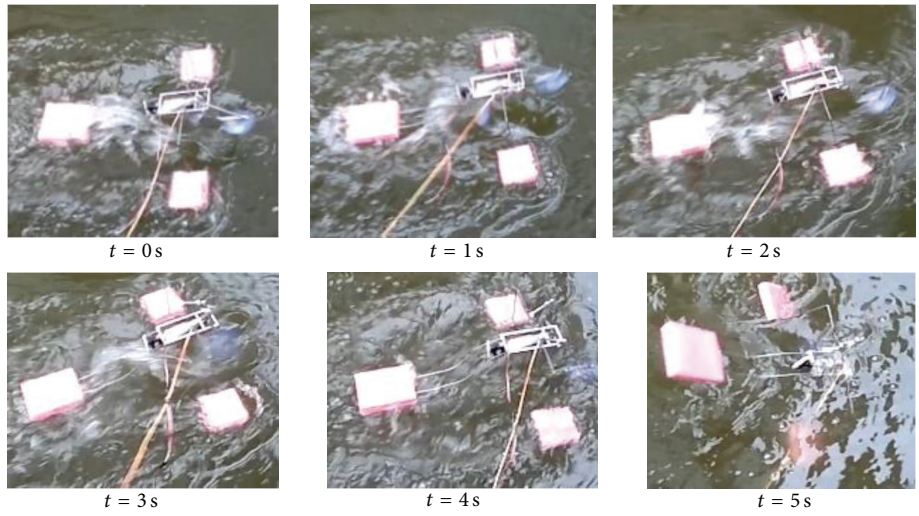


Figure 11: Sequence diagrams of robot walking on water.

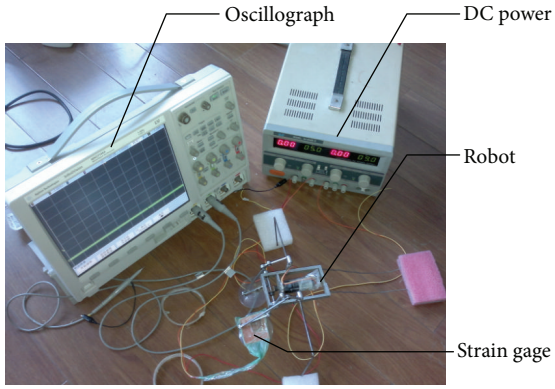


Figure 12: Experimental platform.

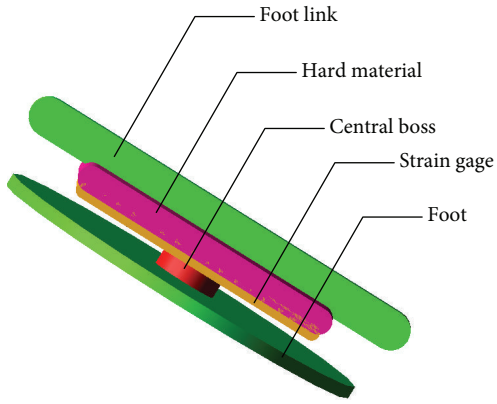


Figure 13: Sketch map for mounting strain gauges.

in Figure 3 of 1.3 N also. The simulation force values are bigger than the realistic ones because the transmission efficiency, the pressure of the atmosphere, and the deformation of the soles are not considered in the simulating process.

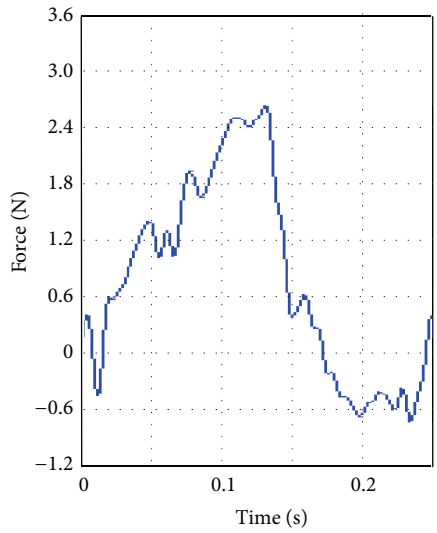


Figure 14: Force curve measured by strain gauge.

6. Conclusions

Inspired by basilisk lizards, a novel biped robot running on water has been designed in this paper, which will offer theoretical and design basis for developing the bionic amphibious biped robot. The virtual prototype of the robot is built by designing the propulsion mechanism with Watt-I planar linkage. At last the prototype of the robot has been made and the water-running experiments have been done. The lifting and propulsion force generated by the propulsion mechanism is about 1.3 N.

While there is still significant work remaining. The motion control method of the amphibious biped mechanism should be studied to implement the control method of the robot motion mode such as going straight, turning, and avoiding obstacle. The composite propulsion mechanism of the amphibious biped robot should be designed, and

the propulsion movement on land and water should be implemented by using the different modes of motion and control of the mechanism.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 50905175) and the National Program on Key Basic Research Project of China (no. 2011CB302106).

References

[1] F. Gao, "Reflection on the current status and development strategy of mechanism research," *Chinese Journal of Mechanical Engineering*, vol. 41, no. 8, pp. 3–17, 2005 (Chinese).

[2] D. Xiao-po and W. Xu-ben, "Development of rescue robot technology and its application in disaster," *Journal of Disaster Prevention and Mitigation Engineering*, vol. 27, no. 1, pp. 112–117, 2007 (Chinese).

[3] Y. S. Song and M. Sitti, "Surface-tension-driven biologically inspired water strider robots: theory and experiments," *IEEE Transactions on Robotics*, vol. 23, no. 3, pp. 578–589, 2007.

[4] S. Floyd and M. Sitti, "Design and development of the lifting and propulsion mechanism for a biologically inspired water runner robot," *IEEE Transactions on Robotics*, vol. 24, no. 3, pp. 698–709, 2008.

[5] J. W. Glasheen and T. A. McMahon, "Vertical water entry of disks at low Froude numbers," *Physics of Fluids*, vol. 8, no. 8, pp. 2078–2083, 1996.

[6] J. W. Glasheen and T. A. McMahon, "Size-dependence of water-running ability in basilisk lizards (Basiliscus basiliscus)," *Journal of Experimental Biology*, vol. 199, no. 12, pp. 2611–2618, 1996.

[7] Z. Wenwei and W. Kejian, *Mechanisms and Machine Theory*, Higher Education Press, Beijing, China, 7th edition, 1997, (Chinese).

[8] W. Xianming, X. Linsen, C. Kai, and L. Minzhou, "Intelligent control method design of foot robot walking on water" (Chinese), *Robot*. In press.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 697415, 11 pages
http://dx.doi.org/10.1155/2013/697415

Research Article

A Hybrid Architecture for Vision-Based Obstacle Avoidance

Mehmet Serdar Güzel[1] and Wan Zakaria[2]

[1] *Department of Computer Engineering, Ankara University, 06830 Ankara, Turkey*
[2] *Faculty of Electrical and Electronic Engineering, UTHM, 86400 Batu Pahat, Johor, Malaysia*

Correspondence should be addressed to Mehmet Serdar Güzel; mguzel@ankara.edu.tr

Received 11 February 2013; Accepted 20 August 2013

Academic Editor: Shao Zili

This paper proposes a new obstacle avoidance method using a single monocular vision camera as the only sensor which is called as *Hybrid Architecture*. This architecture integrates a high performance appearance-based obstacle detection method into an optical flow-based navigation system. The hybrid architecture was designed and implemented to run both methods simultaneously and is able to combine the results of each method using a novel arbitration mechanism. The proposed strategy successfully fused two different vision-based obstacle avoidance methods using this arbitration mechanism in order to permit a safer obstacle avoidance system. Accordingly, to establish the adequacy of the design of the obstacle avoidance system, a series of experiments were conducted. The results demonstrate the characteristics of the proposed architecture, and the results prove that its performance is somewhat better than the conventional optical flow-based architecture. Especially, the robot employing Hybrid Architecture avoids lateral obstacles in a more smooth and robust manner than when using the conventional optical flow-based technique.

1. Introduction

One of the key research problems in mobile robot navigation concerns methods for obstacle avoidance. In order to cope with this problem, most autonomous navigation systems rely on range data for obstacle detection. Ultrasonic sensors, laser rangefinders, and stereo vision techniques are widely used for estimating range. However, all of these have drawbacks. Ultrasonic sensors suffer from poor angular resolution, and laser range finders and stereo vision systems are relatively expensive. Moreover, the computational complexity of stereo vision systems is another key challenge [1]. In addition to their other shortcomings, range sensors are not capable of differentiating between different types of ground surfaces such as pavements and adjacent flat grassy areas. Overall the computational complexity of the avoidance algorithms and the cost of sensors are the most critical factors for real-time applications. The use of monocular vision-based systems can avoid these problems, and they are able to provide appropriate solutions to the obstacle avoidance problem. There are two general types of vision-based obstacle avoidance techniques: those that compute apparent motion and those that rely on the appearance of individual pixels for monocular vision-based obstacle avoidance systems. The first group is called optical flow-based techniques, in which the main idea is to control the robot using optical flow data, from which the heading direction of the observer and time-to-contact values are obtained [2]. One way of using these values is by acting to achieve a certain type of flow. For instance, to maintain ambient orientation, the type of optic flow required is to detect no flow at all. If some flow is detected, then the robot should change the forces produced by its effectors so as to minimize this flow, based on the Law of Control [3]. A second group of techniques is called the appearance-based methods, which in essence rely on qualitative information. They utilize basic image processing techniques which consist of detecting pixels different in appearance from those of the ground and then classifying them as obstacles. The algorithms used perform in real-time, provide a high-resolution obstacle image, and can operate in a variety of environments [4]. The main advantages of these two types of conventional methods are their ease of implementation and ready availability for real-time applications.

2. Literature Survey

Optical flow, as illustrated in Figure 1, is an approximation of the motion field, summarizing the temporal changes in an image sequence. Optical flow estimation is one of the central problems in computer vision. There are several methods which can be employed to determine optical flow, namely, block-based, differential, phase correlation, and variational methods [5, 6]. There has been wide interest in the use of optical flow for vision-based mobile robot navigation. The visual control of motion in flying insects has been shown to provide important clues for navigational tasks such as centred flight in corridors and the estimation of distance travelled, encouraging new biologically inspired approaches to mobile robot navigation using optical flow. Behaviour such as corridor centring, docking, and visual odometry have all been demonstrated in practice using visual motion for the closed loop control of a mobile robot [7]. In recent years, there has been growing amount of literature on optical flow-based mobile robot navigation. Bernardino and Santos-Victor [8] used biologically inspired behaviours based on stereo vision for obstacle detection. A trinocular vision system for mobile robot navigation has been also proposed [9]. These methods, in some ways, emulate corridor following behaviour; nevertheless, their main disadvantage is the need to employ more than one camera. Alternatively, a number of studies relying on monocular vision have proposed the employment of optical flow techniques for mobile robot navigation [2, 4, 10, 11].

Appearance-based methods which identify locations on the basis of sensory similarities are a promising potential solution to mobile robot navigation. One of the main idea behind the strategy is to head the robot towards an obstacle-free position using similarities between the template and the active images [12]. This is called template matching and is discussed in the next section. The similarity between the image patterns can be obtained by using feature detectors, involving corner-based detectors, region-based detectors, and distribution-based descriptors [13]. However, most of these techniques consume a lot of processing time which is not appropriate for real-time systems. In order to handle the performance problem, algorithms are designed based on the appearance of individual pixels. The classification of obstacles is carried out by using differences between pixels in the template and active image patterns, and any pixel that differs in appearance from the ground is classified as an obstacle. The method requires three assumptions that are reasonable for a variety of indoor and outdoor environments, which are as follows.

(i) Obstacles must be different in appearance from the ground;

(ii) the ground must be flat;

(iii) there must be no overhanging obstacles.

The first assumption distinguishes obstacles from the ground, while the second and third assumptions are required to estimate the distances between detected obstacles and the robot. There are several models for representing colour. The main model is the RGB (red, green, and blue) schema which is used in most image file formats; however, colour information in this model is very noisy at low intensity. The RGB format is frequently converted to HSV (hue, saturation, and value) or HIS (hue, intensity, and saturation). Hue is what humans perceive as colour; saturation is determined by a combination of light intensity and the extent to which it is distributed across the spectrum of different wavelengths, and value is related to brightness. In HIS, I is an intensity value with a range from 0 to 1 where 0 represents black and white 1. These colour spaces are assumed to be less sensitive to noise and lighting conditions.

(a) Destination (b) Source (c) Animation

Figure 1: Illustration of flow vectors and motion animation. (a) Destination image, (b) source image, and (c) movement animation.

3. Hybrid Architecture

Optical flow-based methods suffer from two major problems. The first and most important of these is illumination, which is markedly affected by variations in lighting and shadows [3]. Another major issue is sensitivity to noise and distortion. Various integrated methods for solving these problems have been proposed; nevertheless, it is still a key challenge in employing optical flow methodologies for mobile robot navigation. Appearance-based methods have significant processing and performance advantages which make them a good alternative for vision-based obstacle avoidance. Nevertheless, these techniques still suffer from illumination problems and are highly sensitive to floor imperfections, as well as to the physical

structure of the terrain. To overcome these drawbacks, an alternative method has been proposed which essentially relies on a fusion of both techniques, as illustrated in Figure 2. The main strategy behind this proposal is to integrate the results obtained from an appearance-based method into the proposed optical flow-based architecture. In order to achieve this integration, flow equations are updated with respect to an estimated binary image. However, the binary image illustrated with Boolean logic (F/O) needs to be converted into logical expressions in order to be reasoned over. The method used in this study obtains the extreme values (the highest and lowest average magnitude values) from flow clusters. These are subsequently replaced with Boolean values for the binary image in which the highest value is replaced with "O" members and the lowest value is replaced with "F" members. Algorithm 1 illustrates how the estimated Boolean values from the appearance-based method are converted into flow values.

The conversion procedure for each segment can be formalized as follows:

$$c_i = \begin{cases} F_{\max}, & s_i = O, \\ F_{\min}, & s_i = F, \end{cases} \tag{1}$$

where s_i represents the ith segment extracted from the corresponding binary image and c_i is its updated equivalent.

Equation (2) is used to calculate the new heading angle, including the corresponding member of the map. The new heading angle and the updated version of the control equation ($\theta_{u\text{new}}$) can be expressed as follows:

$$\theta_{u\text{new}} = -\left(\frac{\sum |w_{uL}| - \sum |w_{uR}|}{\sum |w_{uL}| + \sum |w_{uR}|} \times n \right), \tag{2}$$

where $\sum |w_{uL}|$ and $\sum |w_{uR}|$ are the sums of the magnitudes of optical flow and converted map regions with respect to the extreme flow values in the visual hemifields on both sides of the robot's body. These can be detailed as follows:

$$\begin{aligned} w_{uL} &= \sum_{i=1}^{n} \left[w_L(i) + w_{AL}(i) \right], \\ w_{uR} &= \sum_{i=n+1}^{2n} \left[w_R(i) + w_{AR}(i) \right], \end{aligned} \tag{3}$$

where w_L and w_R represent the average magnitudes of flow vectors in the left and right clusters, respectively, whereas w_{AL} and w_{AR} represent the converted segments from the binary image (n is the number of clusters and is set to 4).

The flowchart of the overall control architecture is illustrated in Figure 2; the image sequence is used by the optical flow module to calculate flow vectors and corresponding parameters such as focus of expansion (FOE) and time to contact (TTC). Simultaneously, the last obtained image is correlated with a template in order to estimate the free (F) and occupied (O) parts of the current image based on the appearance-based obstacle detection method. The conversion module converts the output of the appearance-based obstacle detection output into flow-based values. The control law is generated based on the inputs provided by both the optical flow and conversion modules (see (2)). Finally, the behaviour module selects the appropriate behavior based on its arbitration mechanism to steer the robot towards a free space.

Figures 3, 4, 5, and 6 present the output of both detection algorithms for different frames captured from a navigation scenario. The control parameters of each frame are included in Table 1.

The hybrid technique has the ability to negotiate and avoid walls and doors by benefiting from the results of the optical flow-based navigation technique using the frontal optic flow to estimate the so-called time to contact before a frontal collision is likely to occur. Furthermore, it possesses the ability to avoid lateral obstacles in both a safer and smoother manner than with the conventional optical flow technique. Figure 5 presents such a scenario where the robot, using the optical flow-based method, is not able to avoid the obstacle, because the system does not generate an appropriate steering angle. The major difficulty with optical flow methods in mobile robot navigation is that despite the assumption of constant illumination, lighting conditions are still vulnerable to environmental factors. This may cause miscalculations of flow vectors. However, the hybrid system integrates the results of the appearance-based method, into the control law which enforces the overall control strategy. For this scenario, the hybrid system generates a sharper avoiding manoeuvre which allows the robot to pass the obstacle without colliding with it. Figure 6 presents another scenario in which the hybrid method generates a safer avoidance manoeuvre when compared with the conventional optical flow method. This is because the hybrid architecture involves merging the optical flow method with the appearance-based method, and this results in a better response to the lateral obstacle. Figures 3 and 6 present the scenarios where the environments are partly open and safe. The results reveal that the control parameters generated by both methods for corresponding scenarios are similar (see Table 1).

Despite their success with lateral obstacles, conventional appearance-based obstacle detection methods tend to fail to detect objects such as walls and doors that span the entire field of view. This is because the appearance-based methods perform a classification of obstacles using differences between pixels in the template and the active image patterns, where any pixel that differs in appearance from the ground is classified as an obstacle. Additionally, region segmentation has some other drawbacks, one of which is that the thresholding technique requires a significant contrast between the background and foreground in order to be successful. This technique essentially works well for the environments which consist of one dominant colour. Accordingly, if the colour of the doors or walls is similar to the floor pattern, the algorithm may easily fail to complete the navigation task. Figure 7 illustrates an example where the appearance-based method is not able to distinguish between the door and the floor in a precise manner due to the similarity in colours of their patterns.

Two additional examples are illustrated in Figures 8 and 9, where the path of the robot is obscured by large obstacles. The

Calculate maximum flow F_{max} and minimum flow F_{min} from the current image
Until the conversion is completed
If the current segment is free
Replace F with F_{min}
else
Replace O with F_{max}
end_until

Algorithm 1: Conversion algorithm.

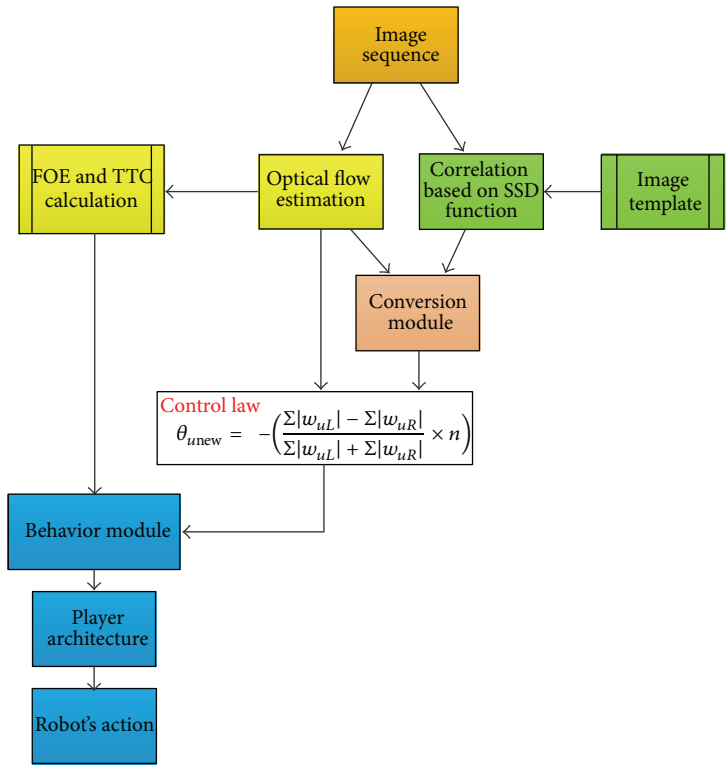
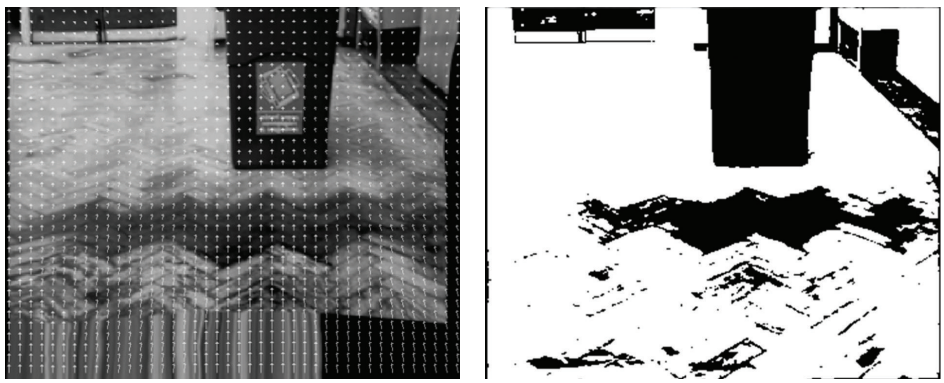
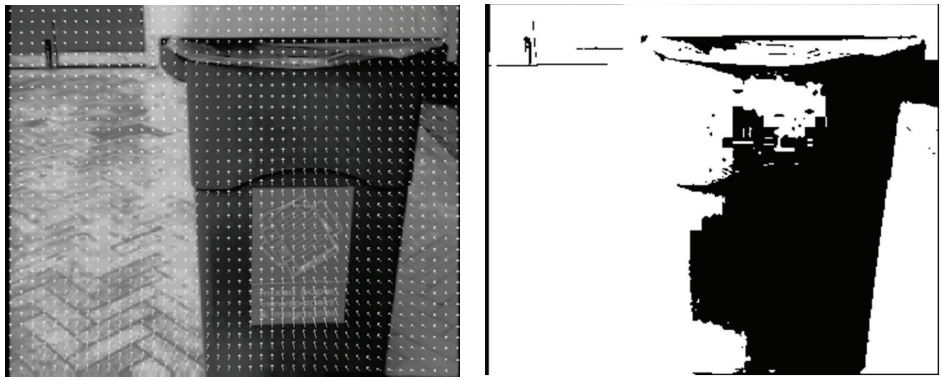


Figure 2: Flowchart of the proposed hybrid architecture.



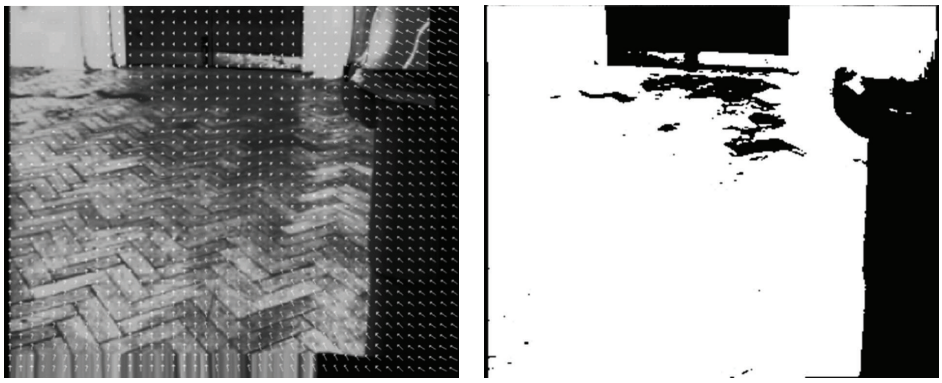
(a) (b)

Figure 3: Frame 1, (a) flow vectors and (b) binary output from appearance-based method.



(a) (b)

Figure 4: Frame 32, (a) flow vectors and (b) binary output from appearance-based method.



(a) (b)

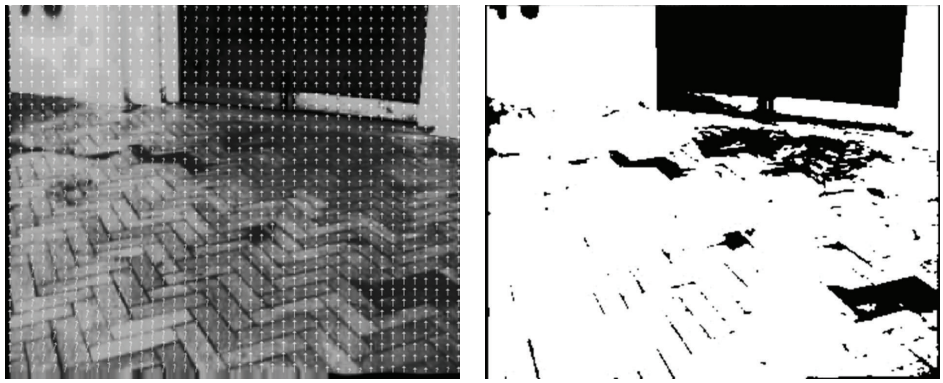
Figure 5: Frame 53, (a) flow vectors and (b) binary output from appearance-based method.

Table 1: Estimated steering angles for experiments.

Left (+)/right (−) Frames	Optical flow w (deg/sec)	Hybrid architecture w (deg/sec)
1	0.63	0.82
32	2.9	5.4
53	6	7.3
97	−0.93	−0.77

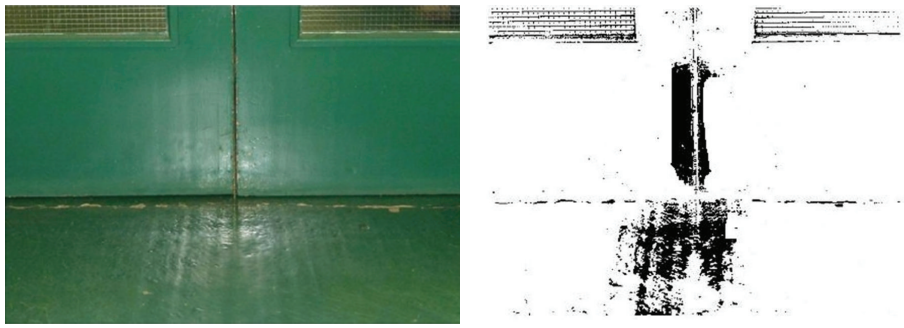
results shown in Figure 8 indicate that both techniques can detect the obstacles. However, due to the lighting conditions, the second technique fails in the segmentation of some parts of the image where reflections are present on the white floor, as shown in Figure 8(b). On the other hand, the first technique estimates the obstacles by successfully using the magnitudes of flow vectors. Figure 9 demonstrates another scenario in which the obstacle is rather close to the goal. Here, the second method is more useful than the conventional optical flow-based technique, despite the extracted stains as shown in Figure 9(b). This is because the appearance-based methods are based on pixel differences which can provide image segmentation independent of distance to the goal. As has been discussed above, the first method focuses on the practical use of optical flow and visual motion information in performing obstacle avoidance task in real indoor environments. However, when the obstacle becomes very close to the robot, the gradients usually cannot be calculated accurately which may result in the incomplete calculation or allocation of flow vectors.

A final example has been added into this study in order to compare the results of the conventional appearance-based detection methods and the proposed method in different illumination conditions. In normal lighting conditions, the appearance-based algorithm is able to separate the given obstacle from the ground surface successfully, as can be seen in Figures 10(a) and 10(b). However, once the illumination conditions are varied in the working environment, the algorithm tends to fail. For instance, in Figure 10(c), the same room becomes brighter due to the change in the illumination condition, and Figure 10(d) exemplifies how the corresponding algorithm fails to segment the image properly. This experiment is a good example of revealing the characteristics of conventional appearance-based approaches, proving that those algorithms are quite vulnerable to change in illumination conditions. Figure 10(f), on the other hand, exemplifies the proposed hybrid algorithm responses in brighter condition which results in estimating flow vectors



(a) (b)

Figure 6: Frame 97, (a) flow vectors and (b) binary output from appearance-based method.



(a) (b)

Figure 7: Similarity of views, (a) original image and (b) binary Image.

successfully and achieves to detect the given obstacle. To evaluate the performance of the proposed navigation method, a series of simulation experiments are discussed in the following section.

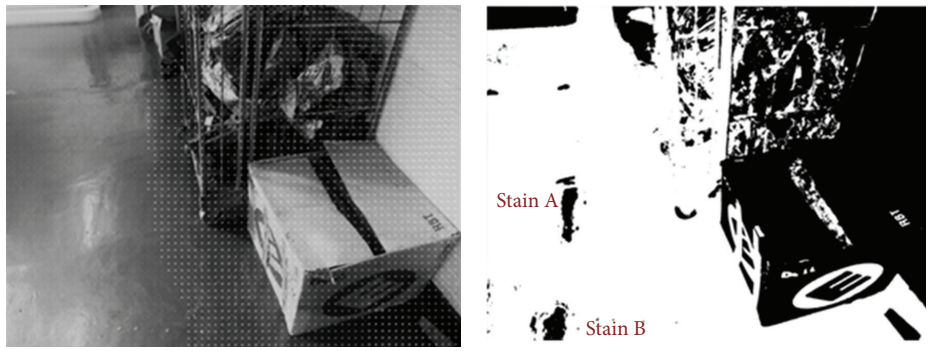
4. Discussion and Conclusions

The navigation systems were uploaded onto the Pioneer mobile robot, shown in Figure 11. All experiments were conducted in and around an area of the Robotics and Automation Laboratory (RAL) at Newcastle University, which has the physical dimensions of 15.60 m × 17.55 m, as illustrated in Figure 11. Hard board panels were used to simulate walls during the experiments.

This section presents the design of the experiments used to evaluate the proposed hybrid vision-based obstacle avoidance system. The experiments were conducted in the test environments shown in Figure 12. In order to verify the performance of the proposed system, the results for each scenario are compared with those of the conventional optical flow method. The main aim of these experiments is to navigate the robot in these environments with regard to the designed scenarios without hitting any obstacles until a certain amount of time has passed. The robot navigates in these experiments at a linear speed of 0.15 m/sec, and the required time limit is 200 seconds to fulfill each scenario. Therefore, once the robot achieves to wander along the environment without colliding until the end of the time limit, it is accepted to complete the task successfully. All overhead lights in the laboratory and corridor environment are turned on during the capture of both snapshot and current images, in an attempt to maintain constant illumination over the entire experimental area. Images were captured at a resolution of 176 × 144 jpg format and then converted to pgm format.

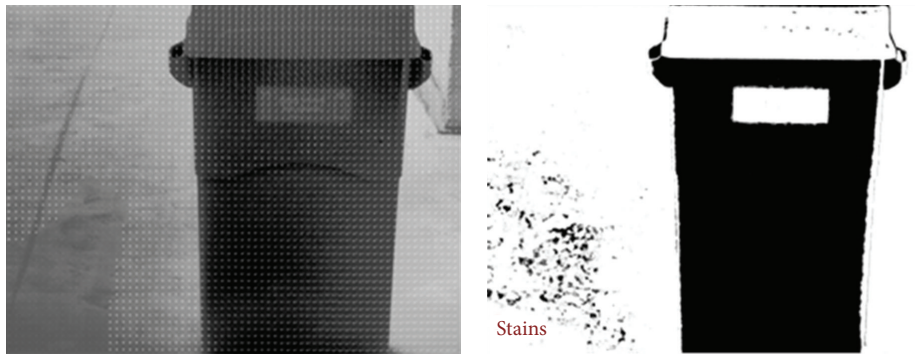
4.1. Definition of Scenarios. Several different scenarios were set up to evaluate the performance of the proposed system. They are arranged in increased level of difficulty. Experiments were conducted in the given test environment as previously discussed.

Each individual test was repeated five times, and the average for each performance parameter was determined. The results for each series of tests were found to be very consistent, principally because the starting position, robot position and feature size, and position of the obstacles are identical for different runs under the same scenario. Two of those different scenarios are discussed in this paper, and an example presenting the limitations of the proposed architecture is demonstrated. Two vision-based obstacle avoidance techniques were employed, namely, the hybrid (FS) and optical flow based (OFB). In order to provide a



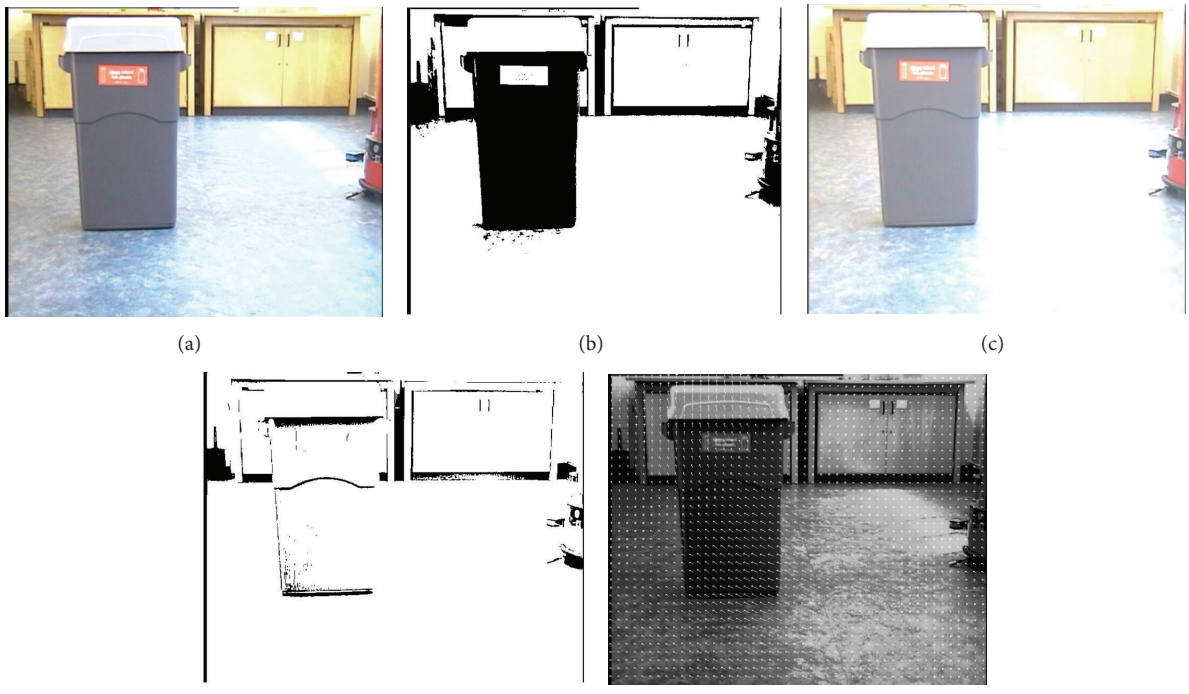
(a) (b)

Figure 8: First large obstacle, (a) original image and (b) binary image.



(a) (b)

Figure 9: Second large obstacle, (a) original image and (b) binary image.



(a) (b) (c)

(d) (e)

Figure 10: Changing illumination, (a) constant illumination and (b). The average of each test was determined binary image, (c) brighter image, (d). The average of each test was determined binary image and (e) flow vectors estimated.

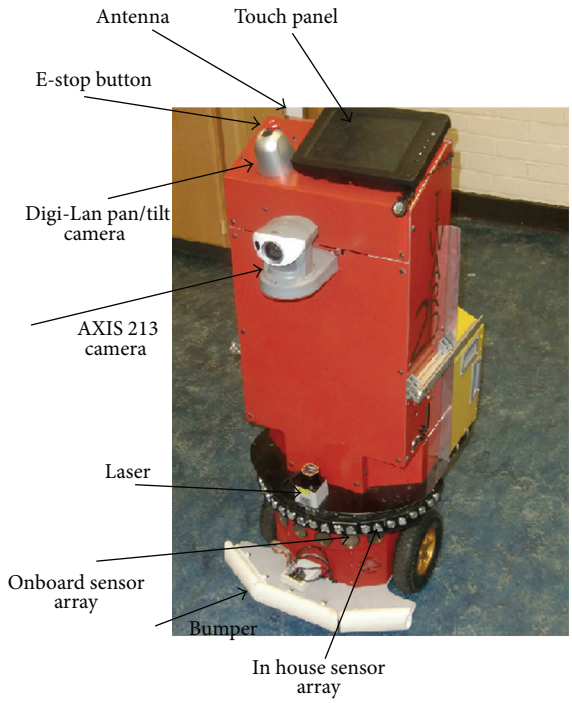


Figure 11: The pioneer robot with additional sensors and peripheral devices.

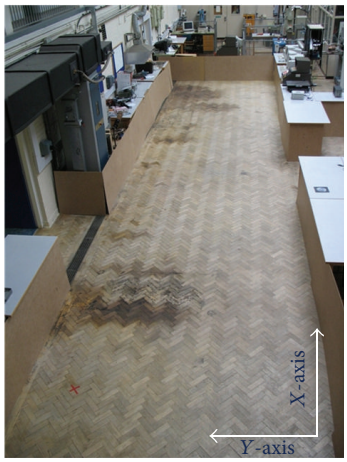


Figure 12: Robotics and Automation Research Laboratory, Newcastle University (including hard-board panels).

precise comparison of the test results, each technique is integrated with the proposed control architecture and the behavioural strategy discussed in Section 3. Table 2 displays the initial parameters used in the navigation algorithms used for conducting the experiments.

Figure 13 presents the first scenario which was conducted in the laboratory environment where the robot was required to navigate in this open environment. The results of the corresponding scenario employing the FS technique are shown in Figure 13(a). The robot navigates through the environment successfully without collision. It negotiates both the door and the wall avoiding them using a 90° left turn

Table 2: Initial parameters for experiments.

Parameters	Descriptions
Initial heading angle	$\theta = 0^{\circ}$
Linear velocity (constant)	$v_c = 0.15$ m/s
Constant turn value	$c_{\text{turn}} = 90^{\circ}$
Maximum range for turning	$n = \pm 20^{\circ}$
Minimum time limit	$T_l = 200$ sec (must move at least 200 sec)

Table 3: Performance measures for scenario 1.

Methods	$\Delta\Omega$ (deg/s^2)	Time (sec)	Collision
Hybrid	4.19	200	No
Optical flow	5.23	182	(1 times)

manoeuvre. It then proceeds to move forward along a left curved trajectory eventually getting back to its start point. The results demonstrate that the FS technique performs smooth and robust behaviour for this navigation task.

The OFB technique performs the navigation without colliding with any obstacle in a smooth manner, as shown in Figure 13(b), so that it negotiates the door and walls, respectively. The results demonstrate that performance in these experiments is surprisingly reliable for this scenario. Table 3 presents the performance measures for each method with this scenario. The FS technique performs the task for each repetition successfully. OFS fails once, but its overall performance is better than expected. Nevertheless it generates a higher value of $\Delta\Omega$ compared to the FS method.

Figure 14 illustrates the second scenario in which the robot is required to navigate in the laboratory environment with three unexpected obstacles placed along its path. Figure 14(a) presents the navigation, results of the FS method for this scenario. The robot begins its navigation, and then it detects the first obstacle. The robot avoids the first obstacle successfully. Subsequently it avoids the second obstacle. After this the robot negotiates walls and the third obstacle, all of which are successfully avoided by following a rectangular path. The navigation results for the OFB technique for this scenario are given in Figure 14(b), where the robot avoids the first obstacle but collides with the second. After this the robot passes the second room and is stopped. Performance measurements of the given scenario are illustrated in Table 5.

4.2. Comparison and Evaluation of Methods. In these test scenarios, the positions of all obstacles in the test environment are unknown to the robot. Two were selected for discussion in this section. The results reveal that the OFB technique addresses the use of optical flow to supervise the navigation of mobile robots. It basically utilizes control laws in aiming to detect the presence of obstacles close to the robot based on information about changes in image brightness. The technique performs better than expected in steering the robot effectively, especially in the open environments as illustrated in Figure 13(b). The major difficulty with employing optical flow in mobile robot navigation is when key information

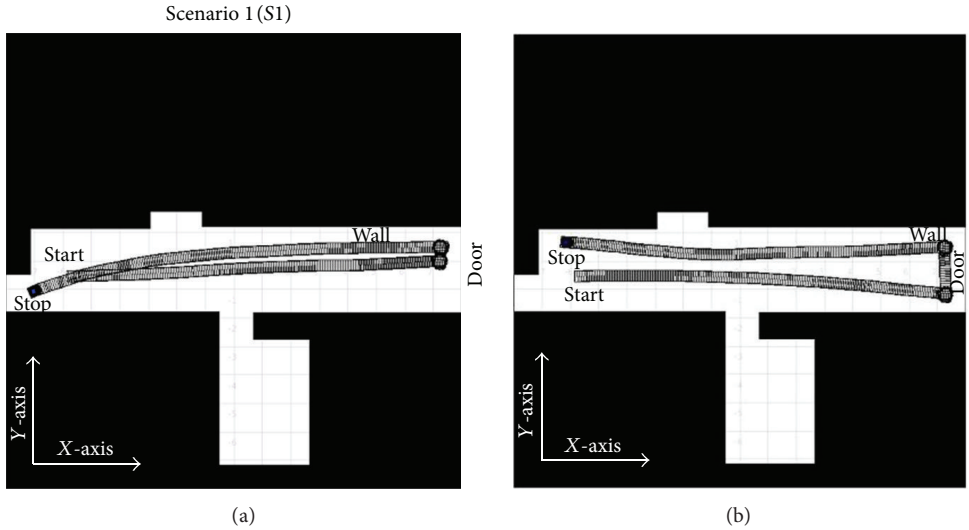


Figure 13: Estimated trajectories for scenario 1, (a) FS and (b) OFB.

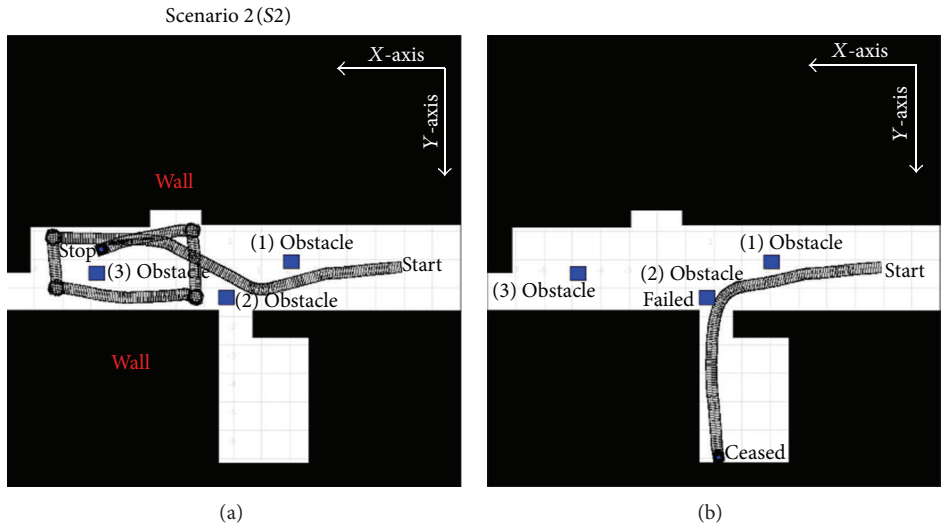


Figure 14: Estimated trajectories for scenario 2, (a) FS and (b) OFB.

is not obtained concerning whether or not motion vectors or changes of illumination change the intensity value of pixels. In addition, despite the assumption of having constant illumination, lighting conditions may significantly change due to environmental factors which optical flow techniques are known to have difficulty in handling. These may cause the miscalculation of flow vectors and can result in collision, as illustrated in Figure 14(b). The OFB is capable of negotiating walls and doors successfully which provides flexibility in this method in partially cluttered indoor environments. However, the OFB technique is not able to avoid external obstacles deliberately located along the path of the robot, as much as the FS technique is able to perform.

It is proposed that the FS technique can improve on the performance of the OFB method, by fusing the results of two techniques in terms of the optical flow-based control law. Figure 14(a) reveals the capacity of this technique in partially cluttered environments, including those with external obstacles. The aim of the technique is to integrate the results of the appearance-based detection technique and optical flow-based navigation architecture. The technique has the ability to negotiate and avoid walls and doors, by benefiting from the results of the optical flow-based navigation technique employing the frontal optic flow to estimate the so-called time to contact before a frontal collision is likely to occur. It is also able to avoid lateral obstacles more smoothly than with the conventional optical flow technique. The outcome of balance strategy tends to maintain equal distances to obstacles on both sides of the robot, exploiting the results of the appearance-based detection technique. The test results

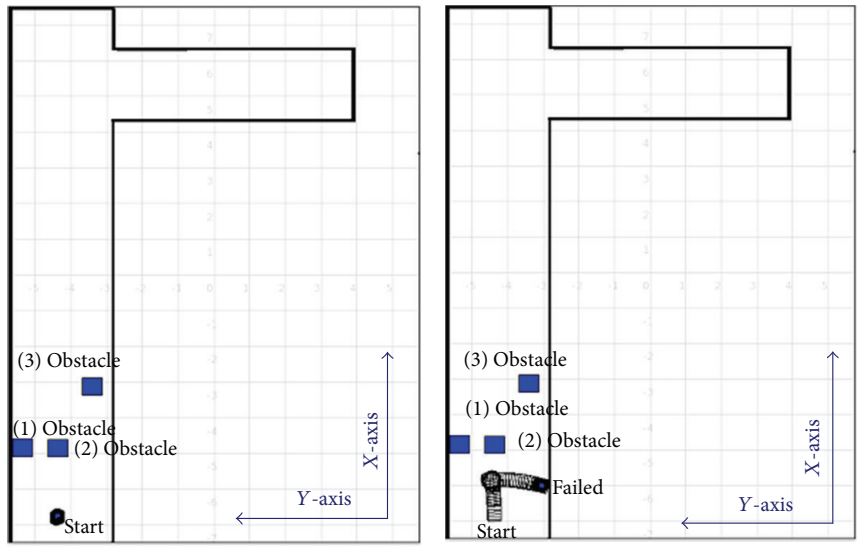


Figure 15: Experimental results for scenario 3 (trap situation), (a) scenario 3 and (b) FS method.

Table 4: Performance measures for scenario 2.

Methods	$\Delta\Omega$ (deg/s^2)	Time (sec)	Collision
Hybrid	6.88	191	(1 times)
Optical flow	7.79	107	(3 times)

Table 5: Performance improvement of the FS over the OFB.

Scenario no.	More consistent	Safer navigation	Longer navigation
S1	20%	25%	9%
S2	26%	100%	78%

reveal that the overall performance of the system is better than that of the conventional technique, but it is still vulnerable to lighting conditions, illumination problems, and floor imperfections.

The characteristics problems of these conventional methods may still affect the performance of the proposed method. Figure 14(b) illustrates the characteristics of the proposed algorithm in cases of frontal obstacles spanning the entire field of view. As the robot remain blind (does not make any decision) and the TTC value indicates the high possibility of collision, the behavioural module is triggered, according to which the *Change Direction* behaviour has a higher priority level than the *Steering* behaviour. Thus, a 90° turning manoeuvre is performed to avoid obstacles. The FS method does not extract the features of the images but only measures the differences between them, which makes the technique appropriate for real-time applications; however, this maybe a disadvantage in more complex situations.

An example illustrating the limitations of the proposed technique is shown in Figure 15 where the robot is not able to avoid both the obstacles and collision with the right wall. This case represents a typical trap situation for this method, where the robot is not able to avoid all obstacles in such complex situations. Table 5 highlights the percentage improvement in performance of the FS over the OFB for the Pioneer robot. Tables 3 and 4 consistently demonstrate improved performance. Accordingly, the FS navigation method offers better overall performance in terms of safety and consistent motion when compared to the OFB method.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

One of the authors sincerely thanks his colleagues who have offered encouragement along this long way, especially all members of the robotics lab.

References

[1] I. Ulrich and I. Nourbakhsh, "Appearance-based obstacle detection with monocular color vision," in *Proceedings of the AAAI National Conference on Artificial Intelligence*, Austin, Tex, USA, July-August 2000.

[2] M. S. Guzel and R. Bicker, "Optical flow based system design for mobile robots," in *Proceedings of the IEEE International Conference on Robotics, Automation and Mechatronics (RAM '10)*, pp. 545–550, Singapore, June 2010.

[3] E. B. Contreras, "A biologically inspired solution for an evolved simulated agent," in *Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference (GECCO '07)*, pp. 206–213, London, UK, July 2007.

[4] G. N. DeSouza and A. C. Kak, "Vision for mobile robot navigation: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237–267, 2002.

[5] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[6] B. Atcheson, W. Heidrich, and I. Ihrke, "An evaluation of optical flow algorithms for background oriented schlieren imaging," *Experiments in Fluids*, vol. 46, no. 3, pp. 467–476, 2009.

[7] D. M. Szenher, *Visual homing in dynamic indoor enviorenments [Ph.D. thesis]*, University of Edinburgh, Edinburgh, UK, 2008.

[8] A. Bernardino and J. Santos-Victor, "Visual behaviours for binocular tracking," *Robotics and Autonomous Systems*, vol. 25, no. 3-4, pp. 137–146, 1998.

[9] A. A. Argyros and F. Bergholm, "Combining central and peripheral vision for reactive robot navigation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, pp. 646–651, Collins, Colo, USA, June 1999.

[10] S. Szabo, D. Coombs, M. Herman, T. Camus, and H. Liu, "A real-time computer vision platform for mobile robot applications," *Real-Time Imaging*, vol. 2, no. 5, pp. 315–327, 1996.

[11] K. Souhila and A. Karim, "Optical flow based robot obstacle avoidance," *International Journal of Advanced Robotic Systems*, vol. 4, no. 1, pp. 13–16, 2007.

[12] R. F. Vassallo, H. J. Schneebeli, and J. Santos-Victor, "Visual servoing and appearance for navigation," *Robotics and Autonomous Systems*, vol. 31, no. 1, pp. 87–97, 2000.

[13] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: a survey," *ACM Computing Surveys*, vol. 38, no. 4, article 13, 2006.

[14] E. B. Contreras, "A biologically inspired solution for an evolved simulated agent," in *Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference (GECCO '07)*, pp. 206–213, London, UK, July 2007.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 175347, 8 pages
http://dx.doi.org/10.1155/2013/175347

Research Article

Study on the Algorithm for Train Operation Adjustment Based on Ordinal Optimization

Yong-jun Chen,[1] Ji-an Yu,[2] Lei-shan Zhou,[3] and Qing Tao[1]

[1] *Beijing University of Civil Engineering and Architecture, Beijing 100044, China*
[2] *Beijing Forestry University, Beijing 100083, China*
[3] *School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Ji-an Yu; yujianbjfu@bjfu.edu.cn

Received 11 April 2013; Accepted 19 July 2013

Academic Editor: Hongxing Wei

It is a crucial and difficult problem in railway transportation dispatch mechanism to automatically compile train operation adjustment (TOA) plan with computer to ensure safe, fast, and punctual running of trains. Based on the proposed model of TOA under the conditions of railway network (RN), we take minimum travel time of train as objective function of optimization, and after fast preliminary evaluation calculation on it, we introduce the theory and method of ordinal optimization (OO) to solve it. This paper discusses in detail the implementation steps of OO algorithm. A practical calculation example of Datong-Qinhuangdao (hereinafter referred to as Da-Qin) heavy haul railway is performed with the proposed algorithm to prove that OO can ensure getting good enough solution with high probability. Particularly, for complex optimization problems with large amount of calculation, OO can greatly increase computational efficiency, and it can save at least one order of magnitude of calculation amount than general heuristic algorithm. Thus, the proposed algorithm can well satisfy the requirements in engineering.

1. Introduction

It is true with practice that a high qualified and efficient TOA system is strongly required with the speeding up of both trains and expansion of train operation density. In China, automation in the railway dispatching and operation system has been gradually adapted into practice at present, which brings up a demand of great significance to research an efficient TOA method adapting to the special characters of the railway and satisfying the requirements of onsite operation. This is no longer a theoretical research but an issue that should to be put into the onsite practice in train dispatching and operation.

TOA mechanism is a high dimensional study of target optimization with nonlinear mixed integer. Many scholars all over the world have investigated algorithms of TOA and have got flourish achievements. The methods these scholars adopted are mainly focused on linear and nonlinear programming, mixed integer programming, branch-bound method, genetic algorithm, Lagrangian relaxation approach, and so forth. For example, in [1], the mechanism of TOA is researched under given train running time and the highest speed; in [2], genetic algorithm is used to establish a satisfaction optimization model of TOA; in [3], the dispatching mechanism is taken as a mixed integer programming problem which could be resolved by branch-bound method. The authors of [4] enumerate all feasible solutions to deal with trains and choose one that minimizes total delaying time of the trains, while the authors of [5] present the earliest conflict optimizing method which overcomes the huge amount of plans in resolving conflicts by combination and is highly efficient in operation, the authors of [6] take minimum total cost on train delay as an objective function and put forward heuristic search function as an accelerated algorithm that in a large degree reduces the number of node graph and in return speeds up the algorithm-solving process, the authors of [7] propose a new method, generating research for the optimization plan automatically through dividing the original problem into many sub-problems and then trying by delamination method to get solutions to those subproblems.

Contributing to hierarchical decision-making method and rolling optimization, the authors of [8] present a general algorithm about both train operation plan and adjustment model.

All of the optimization and algorithms methods (OAM) mentioned above are devoted to getting the best solution to the programming plan; however, the calculating amount grows exponentially with the scale of planning problem growing. A crucial problem is that various mathematic optimization methods are stuck with bottleneck in solving TOA plans in the grand-scale RN, due to unbearably large amount of calculation.

In recent years, simulation optimization and OO methods have become an effective tool in research of discrete event with dynamic systems. As the TOA involves many sectors and factors, it is difficult to propose systematic optimization mathematical models to solve them in practice. It is well known that the problem of train scheduling is a constrained optimization problem and is taken as a nondeterministic polynomial problem (NPP) in mathematics and computer filed. Despite the existence of many intelligent algorithms for solving NPP, such as genetic algorithms, simulated annealing, and ant colony algorithm, they are probabilistic search algorithms and cannot meet real-time requirements as they are time consuming. As a result, the application of OO in solving the problem of TOA emerges as a new way of thinking.

2. Ordinal Optimization (OO)

OO was proposed in 1992 by the Harvard professor Ho et al. [9]. The OO algorithm has been developed into an irreplaceable area of research with hundreds of research papers and is widely used as an important method in resolving simulation optimization problems, especially in those with extraordinarily large space. Below are two main basic ideas of OO.

First, it is easier to compare with "order" than with "value" ("order" versus "value"). This idea could be understood through experiences. Suppose that there are two balls, one in each of your hands. It is easy to tell which ball is heavier just by feeling although we cannot accurately feel the exact weight of each ball. This idea could also be applied in optimizations that are based on simulation. The original system provides a rough model (RM) which shows that, between the two kinds of solutions, we can exactly tell which one is better, although we do not know the exact differences in their performance. If the question is to tell which one is better but not to tell how much it is better than the others, an RM is efficient enough merely by sorting the solution space with an order. For example, in the TOA model (M1), the simulation, with short time or with an average value based on a small number of (or even only one) sample tracks, could be sorted into RMs. It is possible for RMs to save the amount of calculation since it has a great advantage, compared with the accurate model, in spending of time (or economic cost) for calculation. This is the main content of ordinal comparison and also the first basic idea of OO.

Second, we have the idea of goal softening (GS). When the optional solution space is extremely large and it is very difficult or even not feasible to accurately get the best solution, from the perspective of engineering, it is preferable to get a "good enough solution (GES)" as the final result, that is, a "good enough solution" is so acceptable that we do not have to obtain the best one. This idea is especially viable with large-scale optimization problems. What is more, is that we are free to define what a GES in projects is, and a GES can often meet the engineering needs. This is the main content of GS and also the second basic idea of OO.

We can take these two basic ideas in the TOA as follows. When the train delays and the operation of other trains is affected, when there is a plan to increase or decrease one or more operation plans of the trains, or when it is required to adjust the train operation or eliminate conflicts under the condition that malfunction occurred or normal maintenance is ongoing, the train dispatchers do not have to make an adjustment plan after accurately calculating the distance between train and train/section/station, but to change, with sharp and immediate responding, the running order of the trains (ordinal comparison) and adjust the time trains arrive at or depart from the station, trying to find the best arrangement or adjustment under certain constraints. The idea to solve these problems by experiences (GES) is nearly the same with OO. The TOA is a kind of complex optimization problems, its solution space is also very complex because the solution space grows exponentially with the scale of the problems that need to be solved while growing. Calculation for objective function is very complicated and time consuming since the amount of computation is extraordinarily huge, so it could not meet the real-time requirements of TOA. Therefore, OO emerges as a preferable option to obtain "GES" and solve the problems instantly.

Based on these two important features, OO significantly reduces the amount of computation in obtaining solutions that are widely used in engineering and have achieved very pleasant results. This paper discusses how the OO method is applied in the field of TOA and explains in detail the algorithm progress of TOA under complex train network, as well as its implementation steps. At last, with an example of algorithm, this paper proves that the OO method could save more magnitudinous calculation than heuristic algorithm.

3. Model of TOA

TOA of ordinary railway and high-speed railway aims at maximizing trains' arrive-on-time proportion and minimizing actual deviation from train operation plan. Datong-Qinhuangdao Railway (DQR), one of the coal transportation channels in China, whose yearly transportation capacity is large in the world, focused on heavy-haul freight train. Its arrive-on-time proportion was only for reference. The ultimate goal of the adjustment is to raise the average speed of freight trains. For this reason, the least train traveling time is the optimal target of the proposed model. In order to better describe train event and topological structure of station and then to present TOA model (M1).

Consider objective function as

$$\min f_1 = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(q_{ij}\right) + \sum_{i'=1}^{n^s}\sum_{j'=1}^{m^s}\left(q_{i'j'}\right), \tag{M1}$$

that is subject to the following.

Traveling time constraints of train event at station as

$$q_{ij} = \left(y_{ij} - x_{ij} + M\right) \bmod M, \quad i \in [1\cdots n], \quad j \in [1\cdots m]; \tag{1}$$

traveling time constraints of train event at segment as

$$q_{i'j'} = \left(y_{i'j'} - x_{i'j'} + M\right) \bmod M, \quad i' \in [1\cdots n^s], \quad j' \in [1\cdots m^s]; \tag{2}$$

headway constrains between departure times of two same direction consecutive trains as

$$\mathrm{Diff}\left(q_{ij}, q_{i_1 j}\right) \geq \left(1 - A_j^l\right) IF_j^l + A_j^l\left[\mathrm{Len}\left(TL_i^l\right)\right], \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad l \in L; \tag{3}$$

headway constrains between arrival times of the same two direction consecutive trains as

$$\mathrm{Diff}\left(q_{i_1 j}, q_{ij}\right) \geq \left(1 - A_j\right) ID_j^l + A_j\left[\mathrm{Len}\left(TL_i^l\right)\right], \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad l \in L; \tag{4}$$

asynchronous departure-arrival interval times constraints of the same direction train as

$$\left[\left(y_{ij} - x_{i_1 j} + M\right) \bmod M\right] \geq TFD_i^l, \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad l \in L; \tag{5}$$

asynchronous arrival-departure interval time constraints of the same direction train as

$$\left[\left(y_{ij} - x_{i_1 j} + M\right) \bmod M\right] \geq TDF_i^l, \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad l \in L; \tag{6}$$

interval time constraints of two relative direction trains that pass the station at different time as

$$\left[\left(y_{ij} - x_{i_1 j} + M\right) \bmod M\right] \geq TG_i^l, \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad i' \in [1\cdots n^s], \quad j' \in [1\cdots m^s], \quad l \in L; \quad d_{i_1 j} \neq d_{ij}, \quad RI_{i_1' j'}^{s_l} \cap RO_{i'j'}^{s_l} = \phi; \tag{7}$$

headway constraints between asynchronous arrival times of two trains as

$$\mathrm{Diff}\left(q_{ij}, q_{i_1 j}\right) \geq TB_i^l, \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad l \in L; \tag{8}$$

crossing interval time constraints in station of trains as

$$\mathrm{Diff}\left(q_{ij}, q_{i_1 j}\right) \geq TH_i^l, \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad l \in L; \tag{9}$$

free running time, starting additional time and stopping additional time constraints of trains at segment as

$$\mathrm{Len}\left(q_{ij}\right) = TY_j^l + SI_j^l \times TQ_j^l + SO_j^l \times TT_j^l, \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad l \in L; \tag{10}$$

least dwell time constraints of trains in station as

$$\mathrm{Len}\left(q_{ij}\right) \geq TZ_i^l, \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad l \in L; \tag{11}$$

access route and send out route constraints as

$$\left(RI_{i'j'}^{s_l} \cup RO_{i'j'}^{s_l}\right) \cap \left(RI_{i_1' j'}^{s_l} \cup RO_{i_1' j'}^{s_l}\right) \neq \phi, \quad i' \in [1\cdots n^s], \quad j' \in [1\cdots m^s], \quad l \in L; \tag{12}$$

crew working time constraints of trains as

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left(q_{ij}\right) + \sum_{i'=1}^{n^s}\sum_{j'=1}^{m^s}\left(q_{i'j'}\right) \leq TC_{ij}, \quad i \in [1\cdots n], \quad j \in [1\cdots m], \quad i' \in [1\cdots n^s], \quad j' \in [1\cdots m^s], \quad l \in L. \tag{13}$$

The notations of parameters and variables are shown in Table 1.

4. Implementation Steps of the Algorithm

OO used to solve TOA under complex RN can be achieved through the following four steps.

(1) Producing a Set Θ_N for TOA Programs. The prerequisite is not to directly solve algorithm of the TOA model (M1) when applying OO theory to solving the TOA problems but to repeatedly assess the value of objective functions corresponding to all of the feasible solutions within the set Θ_N (a characterization set with a certain number of randomly selected feasible solutions) and determine the pros and cons of each solution. The characterization set Θ_N is composed of a certain number ($N = 100$) of randomly selected feasible solutions taken from the time period when train operation needs to be adjusted. OO helps to transfer operation on solution space to a limited set Θ_N. All of the following steps are working on the representative set Θ_N suppose and that there is no further explanation. When making plans to adjust the train, assume that train 77011 is delayed, the following ones, such as 77013 and 77015, will be influenced to change operation as scheduled, or the three trains will crush into each other in sections S1S2 and S2S3. Consequently, a

Table 1: Subscripts and parameters used in mathematical formulations.

Symbol	Definition
f	Object function of optimization
L	Trains number in RN, $L = \{l \mid l = 1, 2, \ldots, n\}$
V	Set of station, $V = \{v_i \mid i = 1, 2, \ldots, n\}$
V^s	Sets of isolated rail joint, turnout center, intersections of curves, and arrival-departure lines center, $V^s = \{v_i^s \mid i = 1, 2, \ldots, n^s\}$
E^s	Set of the shortest rail that make isolated rail joint, turnout center, intersections of curves, and arrival-departure lines center as vertices, $E^s = \{e_j^s \mid j = 1, 2, \ldots, m^s\}$
B^s	Set of boundary points, $B^s = \{b_j^s \mid j = 1, 2, \ldots, m^s\}$
RI^s	Set of access route that composed by boundary points and line points sequence
RO^s	Set of send-out route that is composed by boundary points and line points sequence
x_{ij}	Starting time for train events $x_{ij} > 0$
y_{ij}	Closing time for train events $y_{ij} > 0$
RI_{ij}^{sl}	Access route for train l at station v_i or at segment e_j
RO_{ij}^{sl}	Send out-route for train l at station v_i or at segment e_j
A_j^l	Driving occlusion mode for train l at segment e_j; automatic block system $A_j^l = 0$ or other $A_j^l = 1$
d_{ij}^l	Direction indicator for train l; $d_{ij}^l = 0$ for an outbound train or $d_{ij}^l = 1$ for an inbound train
SI_j^l	Driving mode for train l enter segment e_j nonstop (stop) $SI_j^l = 0$ or $SI_j^l = 1$
SO_j^l	Driving mode for train l enter segment e_j nonstop (stop) $SO_j^l = 0$ or $SO_j^l = 1$
q_{ij}	Traveling time for train events
IF_j^l	Departure tracing interval time for trains at segment e_j
ID_j^l	Arrival tracing interval time for trains at segment e_j
TY_j^l	Free running time for train l at segment e_j
TQ_j^l	Starting additional time for train l at segment e_j
TT_j^l	Stopping additional time for train l at segment e_j
TB_j^l	Asynchronous arrival interval time for trains at station v_i
TH_j^l	Crossing interval time for trains at station v_i
TL_j^l	Continuous departure interval time for trains at station v_i
TFD_j^l	Asynchronous departure-arrival interval time for trains at station v_i
TDF_j^l	Asynchronous arrival-departure interval time for trains at station v_i
TG_j^l	Asynchronous passing interval time for trains at station v_i
TZ_j^l	Least dwell time standard that train l at station v_i
M	Unit time of diel, $M = 1440$ or $M = 86400$
Len	Length time for train events
Diff	Difference of train events

feasible solution is a solution without any conflicts among trains. Under the premises that train 77011 departures at the scheduled time, there will be several plans to adjust the departure time of the following trains. Every solution from the set Θ_N is not feasible until it passes the feasibility test. We take it as a feasible solution that there is no conflict of trains in the plan and the running time is minimized. If the term "arbitrary adjustment" goes under provision of "adjustment plan," the number of solutions will shrink sharply. For what is shown in Figure 1, OO here is to reduce the size of solution space, transferring operation on all of the adjustment plans (solution space) to certain adjustment plans (finite set Θ_N) occurring in a limited period of time. In the figure, the condition in which conflict points mainly occur is that the time interval that train travels in the section does not satisfy the required interval. In model (M1), the conflict points refer to not only time conflict but also route conflict that the train arrive and depart, and thus the TOAs are to eliminate conflicts both in time and in space. In order to simplify the solving progress of algorithm, this paper discusses the

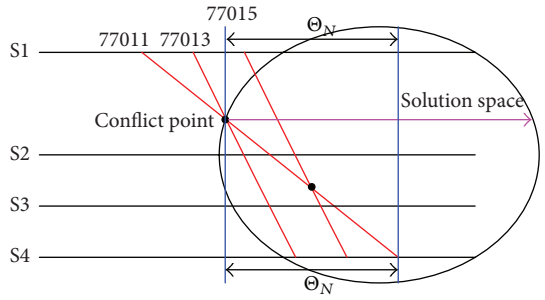


FIGURE 1: Train operation adjustment plan set Θ_N.

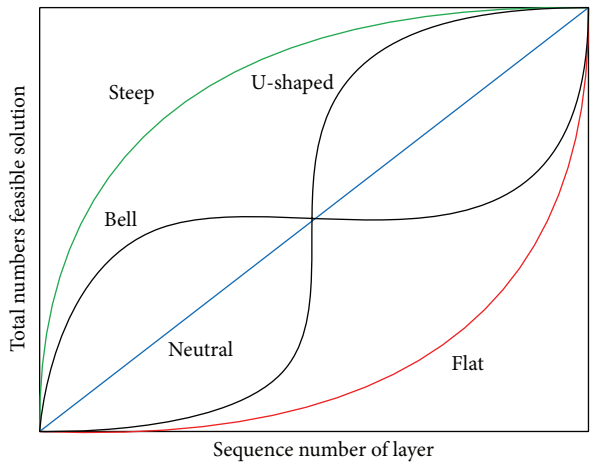


FIGURE 2: Schematic diagram of OPC curve type.

problems only under condition of time conflict, and problems with constraints in both time and space will be discussed in further researches.

(2) Assessing the Ordered Performance Curves (OPC) of TOA Plan, Building up an RM, and Verifying the Type of the TOA Problem. Limited by computing ability, it is unavailable to accurately calculate and get every exact solution of the solution set Θ_N. However, according to the special character of "order comparison" of OO, it is available and feasible to build up an RM for objective function and assess each feasible solution in solution set Θ_N. This is a key point in the theory of OO by roughly assessing the OPC [10] and determinig the type of the problems that need to be researched.

By building up an RM of TOA model (M1), we assess rapidly each of the feasible solutions from set Θ_N based on OO and then sequence them in order from "least" to "most" according to the principle of "the least total travel time of freight trains." As is shown in Figure 2, we take the serial number of each sequenced feasible solution as horizontal axis, the corresponding value of rough assessment as vertical axis, and the curves as OPCs based on the RM of TOA. According to the shape of the OPCs, we can determine which type of the problems belong to, flat style, U-shaped style, neutral style, bell style, or steep style. All types are referring to [10].

Certain requirements need to be satisfied in the RM building for TOA model (M1). If the model is too simple, although the pros and cons of the set Θ_N of all feasible solutions could be preliminarily assessed and the amount of computing OPC shape could be sharply reduced, it could lead to the fact that the results of rough assessment deviate largely from the actual practice and the chosen set S for simulation is too large to reflect the advantages of OO theory; otherwise, if the model is too accurate, the computing amount for solving algorithm in the preliminary assessment of feasible solution set Θ_N will be too large to be accomplished. Therefore, what we need to do is to establish an appropriate RM in order to fully reflect economic advantage of OO in computing. However, the solving algorithm of model (M1) includes not only the departure and arrival time of the train but also the access route and send-out route of trains, which apparently ensure the TOA plan rational and feasible. However, for an RM that not too precise to be calculated, the conditions of station conflicts will not be taken into consideration. According to the special features of constraints in space and time for TOA. By simplifying model (M1) with RM, we get a new model (M2) with only time conflicts, formulated as:

$$\min \quad f_2 = \sum_{i\in\Omega}\left[\sum_{j,j'\in V}\left(y_{ij'} \ominus x_{ij}\right) + \sum_{(j,j')\in E}\left(y_{i(j,j')} \ominus x_{i(j,j')}\right)\right]$$

$$\text{s.t.} \quad x_{ij} \ominus x_{i'j} \geq \left(1 - A^i_{(j,j')}\right) IF^i_{(j,j')} + A^i_{(j,j')}\left[\text{Len}\left(q_{i'j}\right) + TL^i_j\right],$$

$$\forall i \in \Omega, \quad j \in E \cap V, \quad d_{ij} = d_{i'j},$$

$$y_{i'j} \ominus y_{ij} \geq \left(1 - A^i_{(j,j')}\right) ID^i_{(j,j')} + A^i_{(j,j')}\left[\text{Len}\left(q_{ij}\right) + TL^i_j\right],$$

$$\forall i \in \Omega, \quad j \in E \cap V, \quad d_{ij} = d_{j'j},$$

$$y_{ij} \ominus x_{i'j} \geq TDF^i_j \quad i, i' \in \Omega,\ i \neq i', d_{ij} j \in V,\ d_{ij} = d_{i'j},$$

$$y_{ij} \ominus x_{i'j} \geq TG^i_j, \quad i, i' \in \Omega,\ i \neq i',\ j \in V,\ d_{ij} = d_{i'j},$$

$$RI^i_j \cap RO^{i'}_j = \phi,$$

$$y_{ij} \ominus y_{i'j} \geq TB^i_j, \quad i, i' \in \Omega,\ i \neq i',\ j \in V,\ d_{ij} = d_{i'j},$$

$$y_{ij} \ominus y_{i'j} \geq TH^i_j, \quad i, i' \in \Omega,\ i \neq i',\ j \in E,\ d_{ij} = d_{i'j},$$

$$\text{Len}\left(q_{ij}\right) = TY^i_{(j,j')} + SI^i_{(j,j')} \times TQ^i_{(j,j')} + SO^i_{(j,j')} \times TT^i_{(j,j')},$$

$$\forall i \in \Omega,$$

$$\text{Len}\left(q_{ij}\right) \geq TZ^i_j, \quad \forall i \in \Omega,\ \forall j \in V. \tag{M2}$$

Algorithm steps for assessment of the model equation (M2) are as follows.

Step 1. Choose the first train that causes conflicting events in the section of the complex railway network and the section where the train is.

Step 2. If there is no conflicting event, the algorithm is aborted, and the results of the RM of TOA should be outputted.

Step 3. Choose an algorithm section with the earliest time conflict and judge and generate the follow-up moments during which the conflict is eased or solved. Collect all of the moments that satisfy the requirements to ease or solve the conflicts.

Step 4. Choose the minimum travel time from the follow-up moments as the solution of the RM, horizontally traverse the algorithm interval, and ease and solve all of the conflicting points in this section.

Step 5. Take the first moment among the conflict time as the starting point and vertically traverse the second algorithm interval. If there is no conflicting point, then turn to Step 2 or turn to Step 3.

Step 6. Ease or solve all of the conflicting points, and there the algorithm will end, and output a rough TOA plan.

Randomly, we choose 20 adjustment plans. Under the above steps, we assess rapidly the objective function f_2 in model equation (M2) with data from those 20 adjustment plans, calculate the rough values of total travel time (min) of the train, and then sequence those values from "the smallest" to "the largest." All of the feasible solutions can be got and a corresponding OPC will be drafted, as shown in Figure 3. We can judge the optimization type of TOA after comparison with [10]: this OPC belongs to the bell one.

(3) Determine the Set S of TOA That Need to Be Accurately Simulated. OO aims at saving the amount of calculation by building up RMs; that is, its aim is to get a desirable set S of TOA plans by using an RM, with great probability to include at least k solutions that simulate to p layer. It must be ensured to select at least k solutions from the set S of the TOA plan with the least travelling time and without any conflicting events. To highlight the advantages in saving the amount of computation, it is a must that the number of solutions within the set S should be less numerously than that of the whole solution space. After determining the type of the OPC, the value of p and k, and the deviation distribution of the relatively accurate value under rough evaluation—$U \sim [-w, w]$, we can get parameters z, ρ, γ and the regression value of η according to [10]. β, the value of TOA set S, can be calculated by (14) as

$$\beta = \left\lceil e^{z} k^{\rho} p^{\gamma} + \eta \right\rceil. \tag{14}$$

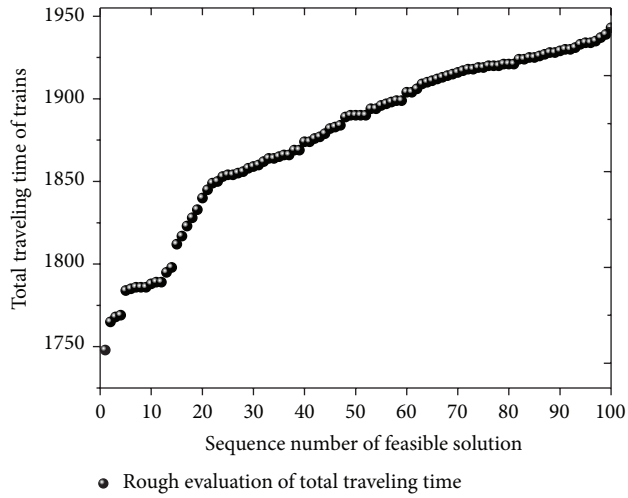


Figure 3: OPC curve of train operation adjustment system.

In (14), $\lfloor e^{z} k^{\rho} p^{\gamma} + \eta \rfloor$ represents the minimum integer which is smaller than or equal to $e^{z} k^{\rho} p^{\gamma} + \eta$.

(4) Choosing a Solution with Least Total Travel Time (TTT) and No Train Conflicts from the TOA Set S. Based on the set S of the TOA plan, we introduce simulation algorithm to solve each subobjective function corresponding to each feasible solution, sequence them from smallest to largest, and select k of them as "GESs." To improve the accuracy of the "GESs," Steps 1–4 and OO calculation should be repeated. After each time of OO calculation, k ones of the "GESs" should be chosen to constitute an alternative set. At last, the best k "GESs" from the alternative set can be chosen as the solution and final plan for TOA.

5. Case Study

The data is from the practical operation of Da-Qin heavy haul railway, with a length of 653 km, 24 CTC control stations, 2 TDCS command centers, and 1 dispatch control center, whose annual carriage capacity lists top in China or even in the world [11]. Thereafter, we use data of Da-Qin Number 1 dispatch control from 14:30 to 18:30, 30 December 2008 to realize simulation based on the above-mentioned algorithm method and steps.

When trains conflict, as shown in Figure 4, firstly, we should find out the first conflict event and define the operation space of TOA and the set Θ_N for the feasible solutions. Secondly, we limit the time space of the feasible solution set Θ_N within the adjustment plan, and then a feasible solution set Θ_N with $N = 100$ is generated.

According to the operation line of Da-Qin heavy haul railway, figure "–○ – ○–" represents combined heavy haul trains and "– ▷ – ▷ –" represents stand-alone trains; besides these, there are still 20 thousand-ton heavy haul trains, C80 unit heavy haul trains, unit heavy haul trains, common goods trains, and so forth, and we distinguish them with different colors and signals.

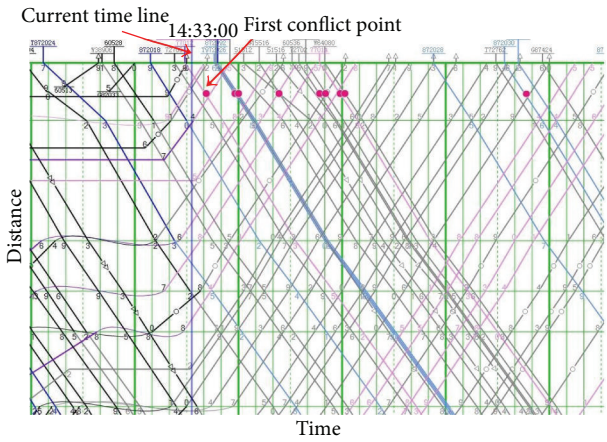


Figure 4: Schematic diagram of conflict point of TOA system.

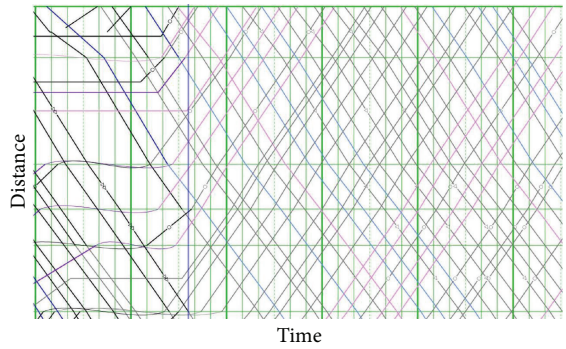


Figure 5: Schematic diagram after untwining conflict automatically by TOA system.

After times of quickly evaluating the model equation (M2), Figure 3 can be got and OPC of the TOA plan can be made—type bell. According to feasible solution set Θ_N and OPC, the set S of TOA plan will be got.

According to [10], there is a specific method to determine set S and further parameter:

(1) Type of OPC: bell.

(2) Range of p: the set Θ_N is accurately evaluated and the top 20%, after sequencing all solutions, are defined as GESs, that is, $p = 100 \times 20\% = 20$.

(3) Range of k: it is required that there is at least one real GES in the selected set S, that is, $k = 1$ and the probability $a\% \geq 95\%$.

(4) Deviation distribution of the relatively accurate value of rough evaluation value f_2—$U \sim [-w, w]$: 30 feasible solutions from set Θ_N are randomly selected, and the total train operation time under both rough evaluation and accurate evaluated value is calculated (i.e., optimized total operation time after adjustment). The difference between the two is used to describe the deviation of rough evaluation to accurate evaluation. by statistic method, we get the standard variance of the deviations, the value is 0.0532; to be more accurate, we make twice the variance of the deviation as the value of w for $U \sim [-w, w]$, that is, $w = 0.1064 \in [-0.5, 0.5]$.

Thus, according to the regression Table 1 of [10], values of the followed will be got: $z = 8.1998$, $\rho = 1.9164$, $\gamma = -2.0250$, and $\eta = 10.00$.

According to (M2), the scale of TOA set S of GESs will be known, that is, 26.1819, which means that 26 feasible solutions from the set Θ_N, after rough evaluation, are selected and then the total train operation time can be accurately calculated.

We make careful simulation for the 26 solutions within set S, calculate their total train time, and sequence the results from the smallest to the biggest. We select the first 5 smallest results as the final GESs of TOA optimizing system, which are 5078, 5095, 5088, 5069, and 5084 minutes. In this case, the final GES is the 86th feasible solution, and total operation time of which is 5069 minutes. The TOA system will automatically work on from the first interval conflict of the descending trains and will solve or optimize all of the conflict points in interval 1, as shown in Figure 5. Identically, the system could gradually solve conflicts in the followed-up intervals quickly layer by layer till all of the conflict points are eased or solved, and an optimized TOA plan of the descending trains with no conflicts will be achieved. The optimized TOA plan of the up-going trains can be achieved by the same way. Then we will have a consummate TOA plan.

All in all, it is an efficacious way to solve the TOA plan based on OO since it is can help to reach the set Θ_N with enough high probability and the real GESs in the whole solution space, which greatly improves the efficiency of calculation in solving problems of TOA. This method used to deal with conflicts in TOA has been verified in the Da-Qin railway as the algorithm could enhance the efficiency and accuracy in train operation, declining greatly the work intensity of the dispatchers.

Under complex conditions of railway networks, based on data from the train dispatch and operation system of Da-Qin heavy haul railway, from the perspective of the real time work, this paper takes minimum travel time of the freight train as objective function of optimization, which aims at adjusting train operation to ease or solve conflict events in intervals and building up a TOA model based on the complex train networks. This method is achieved by simplifying TOA and roughly evaluating the total train operation time of the optimized objective, which could, with greatly decreasing the work in calculation, get GESs for the TOA system. in this paper, "Ordinal Comparison" and "Goal Softening" are made full use of, which avoids directly calculating the TOA model by repeatedly evaluating the value of the objective function. Besides that, to get GESs that meet the requirements, OO could reduce the standards for solving.

As to those optimization problems that are not only real time but also high dimensional and nonlinear with great uncertainty under the complex RN, this method can meet the practical requirements of TOA.

Algorithm analysis and practical application prove well that the method of OO could achieve GESs for TOA plans

while reducing greatly the amount of calculation. In addition, it is supposed to find a standard to judge whether a TOA plan is optimum or not, that is to say, from which perspective the adjustment is the best? In contrast to other regular heuristic methods, OO can well quantize the extent of the pros and cons of the solution results, that is, with what probability TOA plan set S can contain k GESs. These quantitative information is just what regular heuristic methods cannot provide, which is also one of the advantages of applying OO in TOA.

Considering the particularity of Da-Qin heavy haul train, we take the TOA problem in Da-Qin heavy haul rail as a single-objective optimization problem, yet it is not just a single-objective optimization problem but should be a multiple objectives that target at least deviation of TOA plan, minimum train travel time, enhancing the on-time rate of train arrival and departure, and decreasing the number of trains that run late than scheduled. The conditions discussed in the TOA model are limited only to conflicting events in intervals, with considering the conflicts in stations and universality in which the TOA model is applied in different train lines. The questions and problems not mentioned in this paper will be researched and discussed in the future work.

Acknowledgments

This work was financially supported by National Technology Support Program (2009BAG12A10), Research Foundation of State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University (Nos. RCS2009ZT010), and Fundamental Research Funds of Beijing Jiaotong University (2011JBM065).

References

[1] B. Szpigel, "Optimal train scheduling on a single line railway," *Operational Research*, vol. 7, pp. 344–351, 1973.

[2] Y.-R. Chen, Q.-Y. Peng, and Y.-S. Jiang, "Research on a model for adjusting train diagram on double-track railway with satisfactory optimization," *Journal of the China Railway Society*, vol. 25, no. 3, pp. 8–12, 2003.

[3] S. Araya, K. Abe, and K. Fukumori, "An optimal rescheduling for online train traffic control in disturbed situation," in *Proceedings of the 22nd IEEE Conference on Decision and Control*, pp. 489–494, IEEE, San Antonio, Tex, USA, 1983.

[4] R. L. Sauder, "Computer aided train dispatching: decision support through optimization," *Interfaces*, vol. 13, no. 6, pp. 24–37, 1983.

[5] F. Shi, X.-H. Li, J. Qin, and L.-B. Deng, "Earliest conflict optimal method for train operation adjustment on single track railway," *China Railway Science*, vol. 26, no. 1, pp. 106–113, 2005.

[6] D. Jovanovic and P. T. Harker, "Decision support system for train dispatching: an optimization-based methodology," *Transportation Research*, vol. 31, pp. 25–37, 1990.

[7] J. E. Cury, F. A. C. Gomide, and M. J. Mendes, "A methodogy for generation off optimal schedules far an underground technique," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 2, pp. 99–111, 1994.

[8] L.-S. Zhou, "General algorithm and its realization on computer for the train operation adjustment system," *Jouranl of the China Railway Society*, no. 9, pp. 57–65, 1994.

[9] Y. C. Ho, R. S. Sreenivas, and P. Vakili, "Ordinal optimization of DEDS," *Discrete Event Dynamic Systems*, vol. 2, no. 1, pp. 61–88, 1992.

[10] T. W. Edward Lau and Y. C. Ho, "Universal alignment probabilities and subset selection for ordinal optimization," *Journal of Optimization Theory and Applications*, vol. 93, no. 3, pp. 455–489, 1997.

[11] Y.-J. Chen and L.-S. Zhou, "Study on CTC-based heavy haul train dispatch system," *Journal of the China Railway Society*, vol. 30, no. 4, pp. 1–5, 2008.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 234747, 10 pages
http://dx.doi.org/10.1155/2013/234747

Review Article

Autonomous Vehicle Navigation Using Vision and Mapless Strategies: A Survey

Mehmet Serdar Güzel

Computer Engineering Department, Ankara University, 06560 Ankara, Turkey

Correspondence should be addressed to Mehmet Serdar Güzel; mguzel@ankara.edu.tr

Received 11 February 2013; Accepted 25 July 2013

Academic Editor: Shao Zili

This survey addresses the existing state of knowledge related to vision-based mobile robots, especially including their background and history, current trends, and mapless navigation. This paper not only discusses studies relevant to vision-based mobile robot systems but also critically evaluates the methodologies which have been developed and that directly affect such systems.

1. Introduction

A robot is defined as a programmable, self-controlled device. It is in essence a machine which is able to function in place of a living agent [1]. Mobile robots have a long history. Shakey, the world's first mobile robot, was developed in the late 1960s at SRI's Artificial Intelligence Centre (Stanford Research Institute) [1]. Not surprisingly, it has had a substantial influence on present day robotics. Shakey was equipped with various sensors and driven by a problem-solving program called "STRIPS," and used algorithms for perception, world modelling, and actuation. Low-level action procedures were responsible for moving, turning, and path planning tasks. The high-level program could make and execute plans to achieve goals. Another example of an early robot is CART which was developed at Stanford University in 1977 by Hans Moravec as part of his doctoral thesis [2]. However, CART was very slow, not because it was slow-moving by design, but because it was "slow-thinking."

The main reason for this was the difficulty of processing vision data using slow computer processors [3]. Another well-known example was Rover developed at Carnegie Mellon University (CMU) in the early 1980s. Rover was equipped with a camera and was able to perform better than CART. Nevertheless, its thinking and acting were still very slow [4]. In the last decade, the main developments in the area of robotics have come through technological breakthroughs in the areas of computing telecommunications, software, and electronic devices. These technologies have facilitated improvements in intelligent sensors, actuators, and planning and decision-making units which have significantly increased the capabilities of mobile vehicles. The latest trend in robotic intelligence is toward imitating life, for instance in evolutionary robots and emotional control robots.

Another area of technological challenge for the next decade is the development of microrobots and nanorobots for medical applications. On top of this, a paramount challenge will be to find an appropriate balance between human-assisted systems and fully autonomous systems and to integrate technological capabilities with social expectations and requirements [1].

Vision-based indoor mobile robot navigation has been studied for decades and is one of the most powerful and popular sensing method used for autonomous navigation. Compared with other on-board sensing techniques, vision-based approaches to navigation continue to demand attention from the mobile robot research community, due to their ability to provide detailed information about the environment which may not be available using combinations of other types of sensors. The great strides achieved in the area of vision-based navigation systems are significant; however, there is still a long way to go. Two important survey papers that have been published review various aspects of the progress made so far in vision for mobile robot navigation [5, 6].

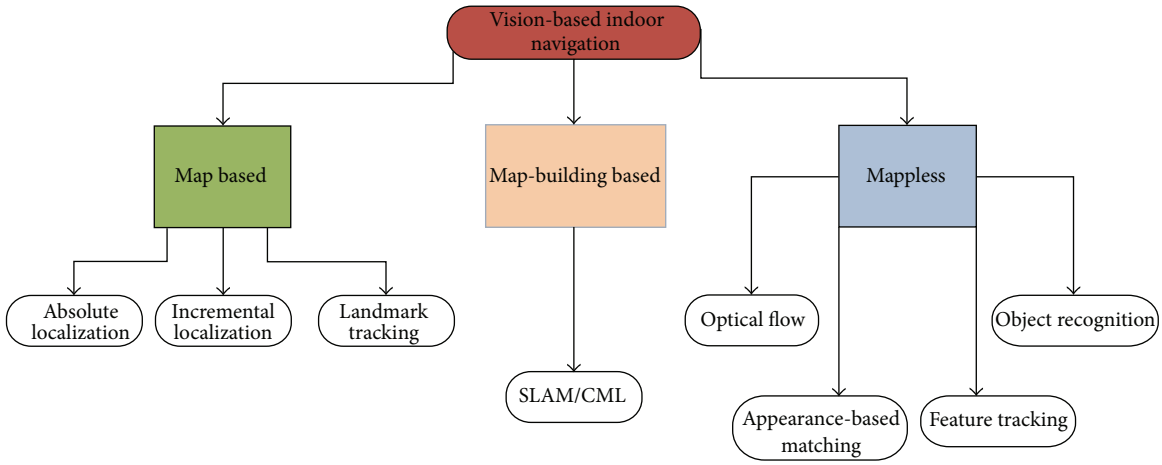


Figure 1: Vision-based indoor mobile robot navigation techniques.

This paper mainly discusses their classification of vision-based navigation systems and attempts to reveal the appropriate state-of-the-art for the indoor environments. A corresponding schema, summarizing state-of-the-art of vision-based navigation, is illustrated in Figure 1. According to the given figure (see Figure 1), there two methods are developed based on map, namely: map-based and map-building-based approaches which will be detailed in the following section.

2. Vision-Based Navigation Using Map

This section aims to detail techniques employing or constructing map to complete navigation problem especially for indoor environments. The first part of this section will introduce map-based navigation approaches in which the robot requires either a metrical or topological map to achieve the given navigation tasks. The second part will highlight the map-building-based methodologies which construct and adapt maps to navigation problems simultaneously. To understand these concepts, a simple analogy was described in [7], and a similar analogy including a daily life scenario is described next. In the scenario it is assumed that a visiting researcher is in Newcastle city centre and needs to return to the Stephenson Building (Newcastle University, Claremont Road) where his office is located. There are various methods that he can follow to reach the office. First, he could memorize the number of steps walked from the university, and he could return by counting the same number of steps. This would be dead-reckoning navigation. He could also buy a map of the environment in order to reach the goal, as in map-based navigation. However, this solution entails that somebody has previously named the streets and have drawn the map. Alternatively, he could also draw his own map using map-building navigation while exploring the city, but this would cost a lot of time and effort. Finally, he could look around trying to find the Claremont Tower and then try to approach it keeping the top of the tower in his field of view (map-less navigation). The goal is to reach the tower, since he knows that Daysh Building is next to it. Autonomous navigation architectures utilize some of these solutions to track a trajectory towards the required goal. Dead-reckoning navigation is the cheapest method and essentially includes an odometry system. However, this solution may include many mechanical problems that produce an increasing error which is unacceptable in long-term navigation. So, an additional perception system is mandatory. Vision is perhaps the most broadly researched perception system.

2.1. Map-Based Navigation. Many techniques employ metric or topological maps to navigate. Navigation techniques need certain knowledge of the environment, and maps may contain different degrees of detail, varying from a complete CAD model of the environment to a simple graph of interconnections between the elements in the environment. One of the key classification criteria in this approach depends on the type of map. For instance, metric-based maps generally favour techniques which produce an optimum, while qualitative methods such as topological maps employ identifiable gateways or landmarks to produce a route.

The main idea behind map-based navigation is essential to provide the robot with a sequence of landmarks expected to be found during navigation, and the task of the vision system is then to search for and recognize the landmarks observed in an image. When the landmarks are recognized, the robot can employ the map to estimate its own position (self-localization) by matching the observation (image) against the expectation (landmark) description in the database. The steps of vision-based localization can be divided into four steps [5].

Acquiring Sensory Information. Acquiring images with associated devices.

Detect Landmarks. Extracting edges, smoothing, filtering, and segmenting regions.

Matching. Identifying landmarks by searching possible matches using the database based on predefined conditions.

Calculate Position. Whenever a set of matches is obtained, the system needs to calculate the position using the observed landmarks' positions in the database.

The start position of the robot is unknown under absolute localization methods. Accordingly, the system must provide exact matching between the current and expected data, derived purely from the entire database. This self-localization problem has been solved either using deterministic triangulation or Monte Carlo-type localization [5, 6]. A detailed implementation of the Monte Carlo localization method to localize a mobile robot without knowledge of its starting location was proposed [8]. Incremental localization assumes that, at the beginning of the navigation, the position of the robot is known approximately. In such cases, the localization algorithm basically keeps track of uncertainties in the robot's position as it executes motion commands and, when the uncertainties exceed a limit uses its sensors for a new fix on its position [5]. The FINALE system is a good example of being able to achieve incremental localization using a geometrical representation of space and a mathematical model of uncertainty regarding the location of the robot [9]. Stereo vision is also preferred by researchers in order to reduce errors [10]. Christensen et al. [11] employed CAD models to represent the working space combined with stereo vision to reduce errors [11].

The final method is landmark tracking in which landmark tracking algorithms determine the position of the robot, detect landmarks on the camera image, and track them in the successive scenes. Landmarks can be artificial or natural. In both cases, the robot needs to recognize the landmarks in order to be able to track them [12–14]. Kabuka and Arenas [12] was the first employing artificial landmark tracking for vision-based navigation. A good example of natural landmark tracking-based navigation was proposed in [13] in which a landmark detection technique that assures stable detection even under variable brightness and an obstacle measurement technique that combines obstacle region segmentation with stereo vision were proposed.

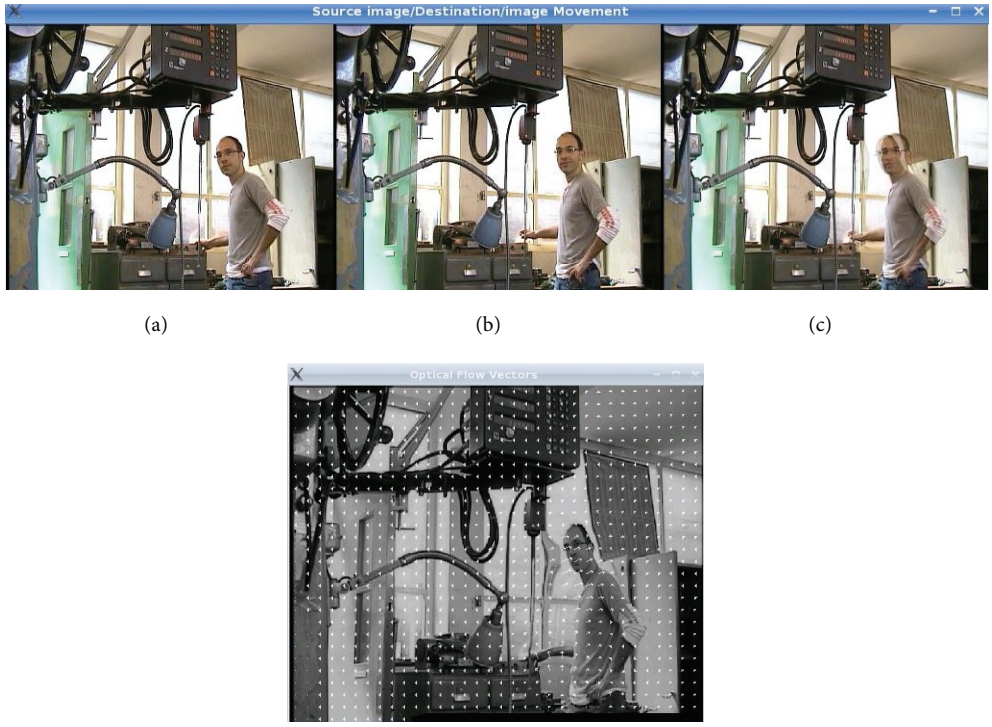
2.2. Map-Building Based Navigation. Sometimes modelling an environment could be difficult particularly if one also has to provide metrical information. An alternative navigation strategy, the map-building-based approach, has been used in both autonomous and semiautonomous systems that entails searching the environment and building a representation of it.

One of the earliest attempts at a map-building technique was carried out by the Stanford CART Robot equipped with a camera. Subsequently, an interest operator algorithm was improved to detect 3D coordinates of the images [15]. The system basically demonstrated the 3D coordinates of the objects, which were stored on a grid having 2 m cells. The map was updated at each iteration; and obstacles were also shown in the map. But the most important problem with the whole system was performance, which took five hours to go 20 metres. Visual navigation studies, employing map-building-based strategies, from the late 1990s to the present, have focused on two methodologies, namely: simultaneous localization and mapping (SLAM) or concurrent mapping and localization (CML) [6]. These methodologies principally propose solutions to automatically overcome the problem of the exploration and mapping of any unknown environment, which essentially entails three simultaneous tasks comprising navigation, mapping, and localization. Vision-based SLAM/CML algorithms mainly employ stereo vision as primary sensor. Se et al. implemented a vision-based mobile robot localization and mapping system in which the robot was equipped with a stereo system to build a 3D map so as to localize simultaneously in 3D [16]. The map was represented as a scale invariant feature transform (SIFT) feature database. It was constantly updated frame by frame and was adaptive to dynamic environments. An alternative and efficient solution to the SLAM problem based on a pair of stereo cameras has also been recently proposed which employs 3D landmarks to localize the robot, as well as constructing an occupancy grid for safe navigation [17]. Other map-building-based navigation techniques are those that impose a human-guided training stage. In such solutions, a human operator guides the robot through an unknown environment. During this process, the robot records images with a stereo camera and constructs the 3D map incrementally. After the map is built, the robot tracks extracted features and computes the optimum path [18].

3. Mapless Navigation

This section discusses a representative selection of mainly reactive visual-based navigation techniques in which navigation is performed without any prior description of the environment. Mapless navigation is scarcely new compared with the previously defined solutions, but new projects using vision systems have been developed in several directions in the last few years. In the systems surveyed in this section, no maps are ever created. The robots can navigate by observing and extracting relevant information about the landmarks in the environment. These elements can be objects such as desks, boxes, and doorways. The mapless visual navigation techniques discussed here are classified in accordance with the main vision technique or types of clues used during the navigation which are optical flow, appearance based, and object recognition navigation techniques based on feature tracking (see Figure 1) [5].

3.1. Optical Flow for Robot Navigation. Optical flow is defined as the motion of all the surface elements from the visual world. When a person moves through the world, the objects in the visual environment flow around this person. The human visual system can detect that person direction of travel from the movement of these objects. Optical flow can be defined as the apparent motion of features in a sequence of images, as shown in Figure 2. It is believed that when the insect is in relative motion with respect to the environment. Accuracy and the range of operation can be altered by changing the relative speed. For instance, features such as "time-to-contact" (depending on speed) are more relevant than distance when it is necessary to avoid an obstacle [5]. Bernardino and Santos-Victor developed an optical flow-based navigation system imitating the visual behavior of bees, called *robee*, and was equipped with a stereo vision system, mimicking the centring reflex behaviour used by a bee



(a) (b) (c)

(d)

Figure 2: Motion estimation, (a) source image, (b) destination image, (c) motion estimation, and (d) estimated flow vectors.

to navigate safely [19]. The robot localizes itself using the difference between the velocity of the image seen with the left eye and the velocity seen in the right eye. If the difference is close to the zero, the robot keeps moving forward. However, if the velocities are different, the robot moves toward the side whose image changes with a lower velocity. Several successful navigation systems have been recently inspired by this centring reflex and implemented to navigate a mobile robot through an unknown indoor environment. For instance, a mobile robot platform utilizing a binocular vision system to estimate optical flow in some way emulates corridor following behaviour to navigate [20]. Duchon et al. implemented a monocular vision-based navigation system based on optical flow algorithms and action modes (behaviours). Simulation and real experiments revealed that the system was capable of navigating in a maze whilst successfully avoiding obstacles [21]. Since then, several optical flow-based control algorithms integrated into a behavioural model have been implemented to be used in real-time applications [22–25].

Additionally, a simple but efficient optical flow-based visual-based navigation system was developed in [26] under the DARPA-Mobile Autonomous Robot Software (MARS) program. The proposed technique first estimates image edge maps by detecting Laplacian of Gaussian (LOG) zero crossings. Afterwards, the edge maps of consecutive frames are used by a patch matching procedure to calculate corresponding optical flow.

3.2. Appearance-Based Methods. Appearance-based methods fundamentally rely on the idea of memorizing the working environment. The main idea is to store images or templates of the environment and associate these images with commands that will steer the robot to its final destination [5]. These methods mainly consist of two procedures. The first one is the training phase in which images or prominent features of the environment are stored as model templates. The models are labeled with certain localization information, which are then associated with appropriate steering command. Secondly, in the navigation stage, the robot has to identify the environment and localize itself by matching the current image with the stored templates [6]. The main problems with the method are to find an appropriate algorithm for the representation of the environment and to define the on-line matching criteria [5, 6].

Beginning in the late 90s, a well-known example of this method was proposed in which the robot merges visual information and their azimuths to build up a representation of its location which is used to estimate the best movement to reach the goal. Another popular technique derived from this approach is the *view-sequenced route* representation which primarily focuses on route construction using a sequence of images and a template matching algorithm to guide robot navigation. Captured images are used to form a sequence of images. Each image in the sequence is associated with the motions required to move to a corresponding destination.

This approach basically introduces the concept of visual memory [27]. Multidimensional histograms provided by the statistical analysis of images are an alternative method to guide mobile robots in appearance-based strategies. Statistical data, including that related to colour, edge density, and texture, are utilized to build a multidimensional histogram database. The recognition of the environment during the navigation stage is achieved by matching the multidimensional histogram of the current image with the multidimensional histogram of the stored templates. This technique consumes less computational resources than when using correlation algorithms [28]. Recent groundbreaking research has proposed an entirely qualitative method in which feature points are automatically detected and tracked throughout the image sequence. In the teaching phase, the (KLT) feature tracker computes the displacement and minimizes the sum of the squared differences between consecutive image frames. The feature coordinates in the replay phase are compared with those computed previously in the teaching phase in order to estimate the steering commands for the robot. Experimental results revealed that the capability of autonomous navigation in both indoor and outdoor environments was successful with the proposed method [29].

An important concept in visual-based mobile robot navigation is the idea of visual homing, inspired by insect behaviour. Insects are able to return to important places in their environment by storing an image of the surroundings while at the goal, and later computing a home direction from a match between this snapshot image and the currently perceived image. For instance, an agent employing a visual homing algorithm captures an image at the home location. It then attempts to return to this location from a nearby position. It compares the current image with the snapshot and infers the direction to the location of the goal from the disparity between these images. It is considered that these aspects of insect behaviour can be a basis for the development of robust navigation algorithms for mobile robots [30].

Visual homing is an appearance-based navigation strategy whose homing algorithms are based on image-based holistic methods using disparities between whole images to compute homing vectors. Image warping is a popular method which is considered to be one of the most reliable visual homing methods for indoor use in this category. It involves calculating the set of all changes in position and orientation between the current and snapshot images. Warping methods distort the current image as if the agent would move according to certain movement parameters. The space of possible movement parameters is then searched for the parameter combination leading to the warped image that is as similar as possible to the stored image. In order to achieve this, warped images are compared to stored image using a pixel-by-pixel correlation measure. The current home vectors are determined based on the strongest similarity between those images [31]. A new and simple visual homing algorithm has been recently proposed where the home position is considered as a charging station [32]. The algorithm basically utilizes the root mean square (RMS) difference and exclusive or (XOR) functions to compare current and snapshot images. The paper reveals that it is possible to implement homing algorithms which allow a robot, fitted with a panoramic camera, return to a reference position from any starting point in an area [32].

3.3. Object Recognition. For the appearance-based approaches previously mentioned, the robot is only able to access few sequences of images that help it to reach its final destination or use predefined images of target goals that it can use to track and pursue. An alternative method has been proposed which essentially employs a symbolic navigation approach instead of memorizing the environment [33, 34].

In this case, the robot utilizes symbolic commands such as "go to the corner in front of you" or "go to the main exit." For instance, a command such as "go to the corner in front" informs the robot that the landmark is a corner and the path points straight ahead. The robot builds a map called an "s-map" which is a 2D grid that stores the projections of the observed landmarks as they are recognized. After the location of the landmark is projected into the s-map, the robot generates the path using GPS-like path planner and employs odometry to reach the goal [5].

3.4. Navigation Techniques Based on Feature Tracking. Tracking the motion of moving elements, including lines, corners, or specific regions in a video sequence has become a popular and robust way of performing navigation. Techniques inspired from tracking are called feature-based approaches which, in essence, determine the trajectory and motion of the robot by tracking and finding relative changes in the position of extracted features [6].

This category can also include feature-based visual homing strategies. Feature-based methods fundamentally segment snapshot and current images into landmarks and background. To operate successfully, feature-based navigation algorithms must extract the same features from snapshot and current images (the feature-extraction problem). Each feature extracted from the *snapshot* image must then be paired with a feature from the *current image* (the correspondence problem). The feature extraction and correspondence problems are difficult to solve in cluttered environments in real-time, since the appearance of landmark changes with viewpoint [31].

One of the earliest studies regarding feature tracking systems introduced a fruit tracking system employing the size and position of a valid fruit's regions in colour images, to control the motion of a fruit-picking robot [35]. Trahanias et al. implemented a robotic system that is able to extract landmarks automatically in indoor environments, using a selective search for landmark patterns which relies both on the workspace and the distinctiveness of the objects in the environment [36]. A KLT tracker-based homing schema was rooted in the extraction of very low-level sensory information, namely, the bearing angles of corners. This was implemented on a robotic platform to evaluate the results [37]. Alternatively, in most cases, feature tracking-based navigation algorithms do not provide an obstacle avoidance module, which must therefore be implemented by other

means depending on the problem. For instance, Li and Yang proposed a behavioral navigation architecture for mobile robots, which utilized a robust visual landmark-recognition system based on genetic algorithms to guide the robot, which used a fuzzy-based obstacle avoidance system and ultrasonic range finder [38].

Two images of the same planar surface in space are related by a homographies. This concept has many practical applications, including mobile robot navigation. In a recent study, Guerrero et al. introduced a method based on homographies computed between current images and images taken in a previous teaching phase with a monocular vision system. The vertical lines (features) were used to estimate the homographies, which are automatically extracted and matched. A complete homography motion can compute rotation and translation up to a scale factor [7]. Besides, a simple bearing only measurement technique relying on monocular vision is proposed with odometry readings in which the reference image and the current image are compared to enhance the localization accuracy of the robot [39]. The scale invariant feature transform (SIFT) method, detailed in Figure 3, is a milestone among techniques to detect the features of images or relevant points, and nowadays it has become a method commonly used in landmark detection applications [40, 41]. The SIFT extracts features that are invariant to image scaling, rotation, and illumination. During the robot navigation process, invariant features which have been detected are then observed from different points of view, angles, and distances and under different illumination conditions and thus become highly appropriate landmarks to be tracked for navigation.

Pons et al. employed the SIFT algorithm for feature-based homing, and are utilized to recover the misalignment of orientation between the current and goal positions. Finally, a home vector between these two positions is calculated using the SIFT matches as a correspondence field [42]. Another original paper using SIFT matching presents a method for search and localization of objects with a mobile robot. The recognition mechanism of the method is based on receptive field cooccurrence histograms and SIFTS matching [43]. Guzel and Bicker have recently proposed a new SIFT-based algorithm which is inspired from a visual homing strategy and adapted to a behavioural architecture to overcome goal-based navigation problems [44].

Another popular robust local feature detector is speeded up robust features (SURF) which is first presented by Bay et al. in 2003. It is mainly inspired by the SIFT descriptor. However, the standard version of SURF is several times faster than SIFT [45]. Pictographs are widely used in a general environment. Accordingly, a novel navigation system has been recently developed for indoor service using SURF algorithm to detect pictographs [46]. A similar study was presented by Hirose et al. in 2011, detecting pictographs with SURF for robot navigation [47].

In an alternative study, SURF algorithm is used to improve the performance of appearance-based localization methods. Experiments were conducted by using several omnidirectional images [48]. Other corresponding studies can be found in references [49, 50].

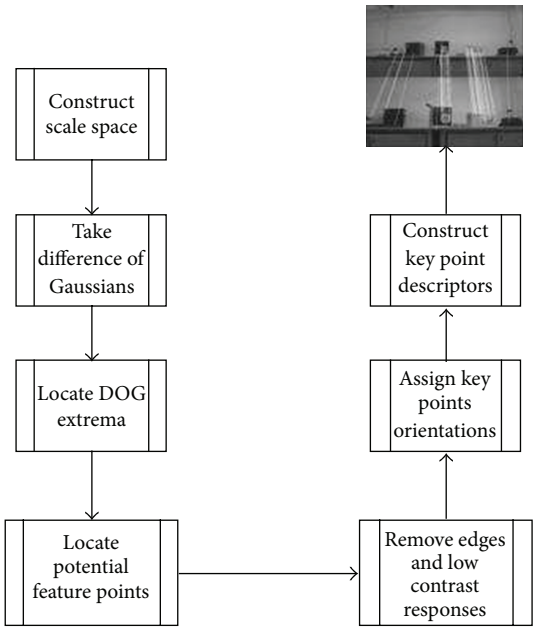


Figure 3: Overview of the SIFT algorithm [39, 40].

A comprehensive research has been recently presented by Martinez-Garcia and Torres [51]. They developed an indoor/outdoor visual odometry scheme for a wheeled mobile robot based on a single passive vision sensor. The authors propose an algorithm which is able to track multiple naturally selected landmarks featured by regional descriptors. The algorithm is in essence an improved version of the maximally stable extremal regions descriptor (MSER). Furthermore, this study was improved by evaluatıng other popular invariant descriptors such as SIFT, Quick SHIFT and feature points extraction algorithm Harris corner detection (HCD) [52], and the fast corner detection (FCD) [53]. Experimental results performed in both indoor and outdoor environments with a mobile robot prove the robustness and consistency of the proposed algorithm. One of the most important criteria in feature tracking-based navigation systems is the computational time consumed by the algorithm. Accordingly, researches have been focused on performance improvement in corresponding algorithms. The most efficient way of performance increment is to parallelize the code or to execute the application on multiprocessor architectures. Various techniques have been proposed to increase parallelism of streaming applications. For instance, Wang et al. have recently presented a comprehensive method in order to reduce schedule length and energy consumption on multiprocessor system-on-chips (MPSoCs). They proposed an ILP design to generate an optimum objective task schedule [54]. Alternatively OpenMP architecture is a robust way of accelerating existing invariant descriptors [44, 55, 56]. The architecture generates multiple threads for parallel programming in a public, shared memory environment.

Visual servoing is another important concept which can be included in this category and is defined as the capability to employ visual information to control the pose of the robot's

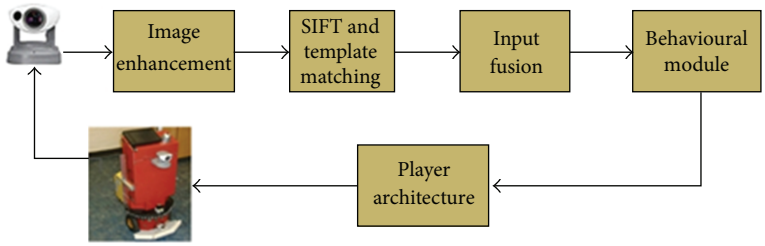


Figure 4: Control architecture of the obstacle avoidance algorithm [64].

end-effectors relative to a target object or a set of target features. The task can also be defined for mobile robots, where it becomes the control of the vehicle's pose with respect to specific landmarks. Thus, Szenher defined the feature-based visual homing algorithms as a type of image-based visual servoing [31]. There are two main approaches for visual servoing systems, namely: position-based visual servoing (PBVS) and image-based visual servoing (IBVS) [57]. PBVS algorithms solve the trajectory problem in workspace. However, in IBVS, the control commands are deduced directly from image features. Kim and Oh have proposed an intelligent mobile robot navigation architecture comprising both of these servoing methods to guide a mobile robot [58]. The IBVS module estimates the motion planning directly from the image space so as to keep the target object always in the field of view. As well as this, the PBVS module is employed to conduct an image-to-workspace transform to plan an optimal pose trajectory directly in the Cartesian space. Another study focuses on docking an AUV to an underwater station with a single standard camera and visual servoing strategy [59].

3.5. Obstacle Avoidance Systems Using Qualitative Information. Obstacle avoidance techniques classified essentially entail extraction of qualitative image characteristics and their interpretation [6]. They are basically defined as sensor-based obstacle avoidance systems which process every item of online sensor data to estimate free and occupied space. These methods are considered useful in avoiding having to compute accurate numerical data such as distance and position coordinates. Lorigo et al. proposed a low resolution vision-based obstacle avoidance architecture consisting of three-dependent vision modules for obstacle detection. These modules were associated with edges, red, green, and blue (RGB) colours, and hue, saturation, and value (HSV) information. The data from these modules was analyzed by a fourth module so as to simultaneously generate motion commands [60]. Ulrich and Nourbakhsh proposed a similar vision-based obstacle avoidance strategy based on monocular vision. The strategy essentially involves assigning each pixel as either obstacle or ground according to its colour appearance [61].

Saitoh et al. integrated this obstacle avoidance technique into a centre followed-based mobile robot navigation architecture. The system does not need prior knowledge of the environment and employs a low cost monocular vision camera as the only sensor needed to navigate the robot safely [62]. The robot has a basic navigation strategy so that it moves towards the centre of the corridor until it encounters an unexpected obstacle. When any obstacle is detected, the robot attempts to avoid it or stops depending on the size of the obstacle. If the robot manages to pass the obstacle successfully, it then localizes itself toward the centre.

ROBOCUP competition has become quite popular and has attracted the attention of many researchers in recent years. The detection of an opponent robot and the ball are two challenging tasks which must be solved efficiently. Fasola and Veloso proposed using image colour segmentation techniques for object detection and grayscale image processing to detect the opponent robots [63]. Guzel and Bicker proposed a new method inspired from feature matching principal and combined Sift-based feature detection and conventional appearance-based method using a fuzzy inference system [64]. The control architecture of the proposed navigation architecture is illustrated in Figure 4.

NASA's twin Mars Exploration Rovers (MERs), Spirit and Opportunity uses a local path planner (GESTALT) equipped with stereo cameras to evaluate terrain safety avoid obstacles [65]. Despite that GESTALT performed well when guiding the rovers around narrow and isolated hazards, it failed when clusters of closely spaced, nontraversable rocks form extended obstacles. To address this problem, a new technology was proposed at the Jet Propulsion Laboratory. According to which Field D^* global path planner has been integrated into rover's software which allows simultaneous local and global planning during navigation. This resulted in better obstacle avoidance and navigation around hazards [66].

A comprehensive review paper about this issue can be found in [67]. The paper introduces major milestones in the development of computer vision for autonomous vehicles over the last decades. Besides, it discusses the design and performance of computer vision algorithms used on Mars in the NASA/JPL Mars Exploration Rover (MER) mission. The paper reveals that the MER mission has been achieved far beyond its primary objectives. The three computer vision algorithms developed for the mission, namely: stereo vision, visual odometry for the rover, and feature tracking for the lander. Despite the limitations of processors deployed, they performed consistently and made significant contributions to the project.

This section details the algorithms and method using mapless strategies for vision-based robot navigation. To summarize this approach, it can be stated that mapless strategies

Table 1: Summary of milestones in visual navigation.

Authors	Category	Strategy	Type of vehicle	Type of sensor
[8]	Map based	Absolute	Ground	Single camera
[9]	Map based	Incremental	Ground	Single camera
[10, 11]	Map based	Incremental	Ground	Stereo vision
[12]	Map based	Landmark	Ground	Stereo vision
[16, 17]	Map building	Slam	Ground	Stereo vision
[19, 20]	Mapless	Optical flow	Ground	Stereo camera
[21, 22]	Mapless	Optical flow	Ground	Single camera
[25, 26]	Mapless	Optical flow	UAV	Single camera
[27, 28]	Mapless	Appearance based	Ground	Single camera
[29, 31]	Mapless	Visual homing	Ground	Omnidirectional
[33, 34]	Mapless	Object recognition	Ground	Single camera
[35, 38, 41, 43, 47, 51]	Mapless	Feature tracking	Ground	Single camera
[48]	Mapless	Feature tracking	Ground	Omnidirectional
[61, 63, 65]	Mapless	Qualitative avoidance	Ground	Single camera
[59]	Mapless	Visual servoing	AUV	Single camera

do not require any explicit representation of the working environment where navigation is to occur.

Principally, these strategies involve recognizing objects found in the environment or tracking those objects using visual clues or observations. Alternatively, motion estimation techniques are widely used in this category. Section two, on the other hand, addresses the strategies using or constructing map for navigation. First, map-based techniques have been introduced which basically employs geometric models or topological maps of the environment. Despite providing fast, robust, and consistent solutions to many problems, they are highly dependent on static maps of working environments which limits the operational capability of the algorithms in this group. Finally, map-building-based strategies have been discussed in which sensors are used to construct geometric models or topological model of the environment. These maps are then used to navigate the robot through corresponding environments. Map-building-based strategies can allow autonomous vehicles to navigate through dynamic environments. However, algorithms used in this group cost a lot of time and effort to obtain a robust model of the environment. Besides, vision-based modelling is challenging, comparing with other sensor systems.

Overall, despite challenges in designing and implementation, mapless strategies developed so far resemble human behaviour more than other approaches and have become applicable to any indoor environment consisting of corridors and doorways.

4. Conclusions

Vision is potentially the most powerful sensing capability in providing reliable and safe navigation. Vision-based navigation techniques are fundamentally classified based on localization techniques. This paper has surveyed each group individually, and as well as vision-based obstacle avoidance techniques are placed under a new section. This paper, inspired from two detailed literature study [5, 6], is essentially the selection of the most outstanding contributions surveyed in the area of vision-based robot navigation. One of the most important differences of this study is that this paper mainly focuses on mapless navigation strategies which are not highlighted and detailed in previously published surveys. Table 1 shows an overview and the most outstanding publications referenced in this survey.

References

[1] P. Nattharith, *Mobile robot navigation using a behavioural control strategy [Ph.D. thesis]*, Newcastle University, Newcastle, UK, 2010.

[2] N. J. Nilsson, *Shakey the Robot*, Artifical Intelligence Center, Computer Science and Techonlogy Division, Standford Research Institute, Menlo Park, Calif, USA, 1984.

[3] T. Hellström, *Biological Foundations of Robot Behavior*, pp. 4–6, Department of Computing Science Umeå University, Umeå, Sweden, 2011.

[4] H. P. Moravec, "The Stanford cart and the CMU rover," *Proceedings of the IEEE*, vol. 71, no. 7, pp. 872–884, 1983.

[5] G. N. DeSouza and A. C. Kak, "Vision for mobile robot navigation: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237–267, 2002.

[6] F. Bonin-Font, A. Ortiz, and G. Oliver, "Visual navigation for mobile robots: a survey," *Journal of Intelligent and Robotic Systems*, vol. 53, no. 3, pp. 263–296, 2008.

[7] J. J. Guerrero, R. Martinez-Cantin, and C. Sagüés, "Visual mapless navigation based on homographies," *Journal of Robotic Systems*, vol. 22, no. 10, pp. 569–581, 2005.

[8] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte Carlo localization for mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '99)*, vol. 2, pp. 1322–1328, May 1999.

[9] A. Kosaka and A. Kak, "Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties," in *Proceedings of the lEEE/RSJ International*

Conference on Intelligent Robots and Systems (IROS '92), July 1992.

[10] L. Matthies and S. A. Shafer, "Error modeling in stereo navigation," *IEEE Journal of Robotics and Automation*, vol. 3, no. 3, pp. 239–248, 1987.

[11] H. I. Christensen, N. O. Kirkeby, S. Kristensen, L. Knudsen, and E. Granum, "Model-driven vision for in-door navigation," *Robotics and Autonomous Systems*, vol. 12, no. 3-4, pp. 199–207, 1994.

[12] M. Hashima, F. Hasegawa, S. Kanda, T. Maruyama, and T. Uchiyama, "Localization and obstacle detection for a robot for carrying food trays," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robot and Systems (IROS '97)*, pp. 345–351, September 1997.

[13] M. R. Kabuka and A. E. Arenas, "Position erification of a mobile robot using standard pattern," *IEEE Journal of Robotics and Automation*, vol. 3, no. 6, pp. 505–516, 1987.

[14] N. Sawasaki, T. Morita, and T. Uchiyama, "Design and implementation of high-speed visual tracking system for real-time motion analysis," in *Proceedings of the IAPR International Conference Pattern Recognition*, vol. 3, pp. 478–483, 1996.

[15] C. E. Thorpe, *An Analysis of Interest Operators for FIDO*, Departmant of Computer Science and The Robotics Institute, Carnegie-Mellon University, Pittsburgh, Pa, USA, 1984.

[16] S. Se, D. Lowe, and J. Little, "Vision-based mobile robot localization and mapping using scale-invariant features," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '01)*, May 2001.

[17] K. Kidono, J. Miura, and Y. Shirai, "Autonomous visual navigation of a mobile robot using a human-guided experience," *Robotics and Autonomous Systems*, vol. 40, no. 2-3, pp. 124–132, 2002.

[18] R. Sim and J. J. Little, "Autonomous vision-based robotic exploration and mapping using hybrid maps and particle filters," *Image and Vision Computing*, vol. 27, no. 1-2, pp. 167–177, 2009.

[19] A. Bernardino and J. Santos-Victor, "Visual behaviours for binocular tracking," *Robotics and Autonomous Systems*, vol. 25, no. 3-4, pp. 137–146, 1998.

[20] A. P. Duchon, W. H. Warren, and L. P. Kaelbling, "Ecological robotics," *Adaptive Behavior*, vol. 6, pp. 1–30, 1994.

[21] S. Temizer, *Optical flow based local navigation [M.S. thesis]*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 2001.

[22] C. McCarthy and N. Barnes, "Performance of optical flow techniques for indoor navigation with a mobile robot," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '04)*, vol. 5, pp. 5093–5098, usa, April 2004.

[23] K. Souhila and A. Karim, "Optical flow based robot obstacle avoidance," *International Journal of Advanced Robotic Systems*, vol. 4, no. 1, pp. 13–16, 2007.

[24] M. S. Guzel and R. Bicker, "Optical flow based system design for mobile robots," in *Proceedings of the IEEE International Conference on Robotics, Automation and Mechatronics (RAM '10)*, pp. 545–550, Singapore, June 2010.

[25] B. Herissé, T. Hamel, R. Mahony, and F. Russotto, "Landing a VTOL unmanned aerial vehicle on a moving platform using optical flow," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 77–89, 2012.

[26] S. Temizer and L. P. Kaelbling, *Optical Flow Based Local Navigation*, MIT, Cambridge, Mass, USA, 2003.

[27] P. Gaussier, C. Joulain, S. Zrehen, J. P. Banquet, and A. Revel, "Visual navigation in an open environment without map," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robot and Systems (IROS '97)*, pp. 545–550, September 1998.

[28] Y. Matsumoto, M. Inaba, and H. Inoue, "Visual navigation using view-sequenced route representation," in *Proceedings of the 13th IEEE International Conference on Robotics and Automation (ICRA '96)*, pp. 83–88, April 1996.

[29] C. Zhou, Y. Wei, and T. Tan, "Mobile robot self-localization based on global visual appearance features," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '03)*, pp. 1271–1276, September 2003.

[30] C. Zhichao and S. T. Birchfield, "Qualitative vision-based mobile robot navigation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '06)*, pp. 2686–2692, May 2006.

[31] D. M. Szenher, *Visual homing in dynamic indoor enviorenments [Ph.D. thesis]*, University of Edinburgh, Edinburgh, UK, 2008.

[32] P. Arena, S. de Fiore, L. Fortuna, L. Nicolosi, L. Patané, and G. Vagliasindi, "Visual homing: experimental results on an autonomous robot," in *Proceedings of the 18th European Conference on Circuit Theory and Design 2007 (ECCTD '07)*, pp. 304–307, August 2007.

[33] D. Kim and R. Nevatia, "Recognition and localization of generic objects for indoor navigation using functionality," *Image and Vision Computing*, vol. 16, no. 11, pp. 729–743, 1998.

[34] D. Kim and R. Nevatia, "Symbolic navigation with a generic map," *Autonomous Robots*, vol. 6, no. 1, pp. 69–88, 1999.

[35] R. C. Harrell, D. C. Slaughter, and P. D. Adsit, "A fruit-tracking system for robotic harvesting," *Machine Vision and Applications*, vol. 2, no. 2, pp. 69–80, 1989.

[36] P. E. Trahanias, S. Velissaris, and T. Garavelos, "Visual landmark extraction and recognition for autonomous robot navigation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robot and Systems (IROS '97)*, pp. 1036–1042, September 1998.

[37] A. A. Argyros and F. Bergholm, "Combining central and peripheral vision for reactive robot navigation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, pp. 646–651, June 1999.

[38] H. Li and S. X. Yang, "A behavior-based mobile robot with a visual landmark-recognition system," *IEEE/ASME Transactions on Mechatronics*, vol. 8, no. 3, pp. 390–400, 2003.

[39] M. C. Deans, *Bearings-only localization and mapping [Ph.D. thesis]*, Mellon University, Pittsburgh, Pa, USA, 2005.

[40] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, pp. 1150–1157, IEEE, Kerkyra, Greece, September 1999.

[41] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[42] J. S. Pons, W. Hübner, H. Dahmen, and H. A. Mallot, "Vision-based robot homing in dynamic environments," in *Proceedings of the 13th IASTED International Conference on Robotics and Applications*, pp. 293–298, ACTA Press, Wuerzburg, Germany, August 2007.

[43] K. Sjöö, G. D. Lopez, C. Paul, P. Jensfelt, and D. Kragic, "Object search and localization for an indoor mobile robot," *Journal of Computing and Information Technology*, vol. 17, no. 1, pp. 1–12, 2009.

[44] M. S. Guzel and R. Bicker, "A behaviour-based architecture for mapless navigation using vision," *International Journal of Advanced Robotic Systems*, vol. 9, pp. 1–13, 2012.

[45] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2003.

[46] D. Chugo, K. Hirose, K. Nakashima, S. Yokota, H. Kobayashi, and H. Hashimoto, "Camera-based navigation for service robots using pictographs on the crossing point," in *Proceedings of the 38th Annual Conference on IEEE Industrial Electronics Society*, Monteral, Canada, October 2012.

[47] K. Hirose, D. Chugo, S. Yokota, and K. Takase, "Service robots navigation using pictographs detection for indoor environment," in *Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society (IECON '11)*, pp. 2170–2175, November 2011.

[48] T. Tasaki, S. Tokura, T. Sonoura, F. Ozaki, and N. Matsuhira, "Mobile robot self-localization based on tracked scale and rotation invariant feature points by using an omnidirectional camera," in *Proceedings of the lEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '10)*, October 2010.

[49] K. Wetzelsberger, T. Pfannschmidt, and T. Ihme, "Real-time image processing for motion planning based on realistic sensor data," in *Proceedings of the 8th IEEE International Workshop on Robotic and Sensors Environments (ROSE '10)*, pp. 158–163, October 2010.

[50] Y. Hagiwara, T. Shoji, and H. Imamura, "Position and rotation estimation for mobile robots in outside of recording path using ego-motion," *IEEJ Transactions on Electronics, Information and Systems*, vol. 133, no. 2, pp. 356–364, 2013.

[51] E. Martinez-Garcia and D. Torres, "Robot visual odometry by tracking multiple natural landmarks with invariant descriptors," ROSSUM, 2011.

[52] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151, August 1988.

[53] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1508–1511, October 2005.

[54] Y. Wang, D. Liu, Z. Qin, and Z. Shao, "Optimally removing intercore communication overhead for streaming applications on MPSoCs'computers," *IEEE Transactions*, vol. 62, no. 2, pp. 336–350, 2013.

[55] S. Warn, W. Emeneker, J. Cothren, and A. Apon, "Accelerating SIFT on parallel architectures," in *Proceedings of the IEEE International Conference on Cluster Computing and Workshops (CLUSTER '09)*, pp. 1, 4, September 2009.

[56] S. Jia, X. Yin, and X. Li, "Mobile robot parallel PF-SLAM based on OpenMP," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '12)*, pp. 508–513, 2012.

[57] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, 1996.

[58] S. Kim and S. Oh, "Hybrid position and image based visual servoing for mobile robots," *Journal of Intelligent and Fuzzy Systems*, vol. 18, no. 1, pp. 73–82, 2007.

[59] P. Lee, B. Jeon, and S. Kim, "Visual servoing for underwater docking of an autonomous underwater vehicle with one camera," in *Proceedings of the OCEANS 2003*, vol. 2, pp. 677–682, San Diego, Calif, USA, September 2003.

[60] L. M. Lorigo, R. A. Brooks, and W. E. L. Crimson, "Visually-guided obstacle avoidance in unstructured environments," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robot and Systems (IROS '97)*, vol. 1, pp. 373–379, September 1997.

[61] I. Ulrich and I. Nourbakhsh, "Appearance-based obstacle detection with monocular color vision," in *Proceedings of the AAAI NationalConference on Artificial Intelligence*, September 2000.

[62] T. Saitoh, N. Tada, and R. Konishi, "Indoor mobile robot navigation by central following based on monocular vision," *IEEJ Transactions on Electronics, Information and Systems*, vol. 129, no. 8, pp. 1576–1584, 2009.

[63] J. Fasola and M. Veloso, "Real-time object detection using segmented and grayscale images," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '06)*, pp. 4088–4093, May 2006.

[64] M. S. Guzel and R. Bicker, "Vision based obstacle avoidance techniques," in *Recent Advances in Mobile Robotics*, pp. 83–108, InTech, New York, NY, USA, 2012.

[65] J. Carsten, A. Rankin, D. Ferguson, and A. Stentz, "Global path planning on board the mars exploration rovers," in *Proceedings of the 2007 IEEE Aerospace Conference*, March 2007.

[66] J. Carsten, A. Rankin, D. Ferguson, and A. Stentz, "Global planning on the mars exploration rovers: software integration and surface testing," *Journal of Field Robotics*, vol. 26, no. 4, pp. 337–357, 2009.

[67] L. Matthias, M. Maimone, and A. Johnson, "Computer vision on Mars," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 67–92, 2007.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 484710, 8 pages
http://dx.doi.org/10.1155/2013/484710

Research Article

Research on the Detection of Metal Debris with Microplane Inductance Sensor

Bendong Liu, Fangzhou Zhang, Yude Wu, and Desheng Li

College of Mechanical Engineering and Applied Electronics Technology, Beijing University of Technology, Beijing 100124, China

Correspondence should be addressed to Bendong Liu; liubendong@bjut.edu.cn

Received 22 March 2013; Accepted 16 July 2013

Academic Editor: Shao Zili

The debris detection system is simulated and analyzed with the software of Maxwell 14 in this paper. The magnetic induction intensity and the magnetic density of the detection system with metal debris are simulated. The static experimental system is designed to measure the inductance caused by different metal debris. The simulation and experimental result indicate that the nonferromagnetic metal debris reduces the inductance of microplane inductance sensor and that ferromagnetic metal debris increases the inductance of microplane sensor. The detection of metal debris with microplane sensor is feasibly proved by the research. This paper provides a model for detecting the debris with a plane eddy current sensor and a case for the 3D simulation of the eddy current. This work may have some significance for improving the efficiency of the plane eddy current sensor.

1. Introduction

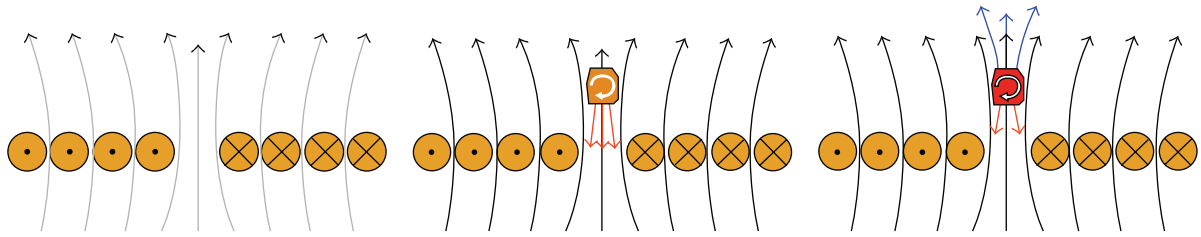
Metal debris will be produced by the friction of the mechanical equipment during the running and manufacturing. It will suspend in oil of the equipment system and then do harm to the equipment. Therefore, the detection of the debris is important for predicting the failure of the wear, reducing major accidents and economic losses [1]. There are many methods that can be selected to detect the debris in the oil. The detection with microinductance has been studied by many scholars because of the high sensitivity, easy integration, and modularization. It has been applied to the measurement of displacement, speed and nondestructive testing [2]. Feng et al. [3] researched the inductance, and quality factor of the plane inductor, and concluded that the inductor had a good inductance and quality factor while the frequency was 1 MHz. Wang et al. [4] studied the plane inductance with the software of Maxwell. Gao et al. [5] and Sadler et al. [6] calculated the inductance of the microplane inductance. Fan et al. [7] researched the metal debris detection with solenoid inductance sensor. Du et al. designed an experiment for detecting the debris in a microchannel with a microplane sensor, which could distinguish the different sizes of the debris by the change of the sensor inductance [8, 9]. He also studied the double microplane inductance sensor [10]. However, most of these researches did not deal with the quantitative influence of the magnetic field and the change of the inductance in calculation affected by the metal debris. The model of debris detecting system is built with the software of Maxwell 14 in this paper in order to simulate the changes of inductance and the changes of magnetic field of the coil. Experimental detection of metal debris is also carried on. The results of simulation and experimental are compared.

2. Detection Principle and the Design of Detection System

2.1. Detection Principle. The principle is that different metal debris has different influence on the inductance of the microcoil. The ferromagnetic debris and nonferromagnetic metal debris can be distinguished according to the influence on the inductance of the microcoil. The size of debris can also be recognized according to the quantity of changes on the inductance. The inductance of microcoil can be calculated with formula (1):

$$L = \left(\frac{1}{i^2}\right) \int \Phi di. \tag{1}$$

Here Φ is the magnetic flux and i is the current element.



(a) The coil with no debris (b) The coil with nonferromagnetic debris (c) The coil with ferromagnetic debris

Figure 1: The schematic drawing of the microinductance sensor. The black arrow lines indicate the magnetic lines generated by the coil, the orange arrow lines indicate the magnetic lines of force generated by eddy current, the blue arrow lines indicate the increasing magnetic lines by the ferromagnetic debris, and the white arrow lines indicate the direction of the eddy current, the irregular geometry indicates the metal debris, and the circular arrangement indicates the coil.

The magnetic flux Φ can be calculated by formula (2)

$$\Phi = \int Bds. \tag{2}$$

Here B is the magnetic induction, s is the area element.

There will be an alternating magnetic field around microcoil while the alternating current passes the microinduction coil. The relation between the magnetic field and the current is shown in Figure 1(a). The magnetic field will be affected while nonferromagnetic metal debris is near to the microcoil as shown in Figure 1(b). The eddy current will be generated in the metal debris because of the alternating magnetic field [11]. The direction of the magnetic field that generated by eddy current is opposite to the magnetic field that generated by the microcoil. So the result is that the magnetic field of the microcoil decreases. It can be indicated that the inductance of microcoil will decrease while the magnetic field reduces with formulas (1) and (2). Therefore, the inductance will decrease while the nonferromagnetic metal debris nears the sensor. The magnetic field of the microcoil with ferromagnetic metal debris is shown in Figure 1(c). The ferromagnetic metal has large permeability. So the magnetic induction will increase due to the magnetization effect. The magnetization effect is greater than eddy current effect. The inductance of microcoil will increase because of the large permeability [12]. Therefore, the material of the metal debris can be recognized according to the increase or decrease of inductance. The debris's size can be identified with the quantity of inductance change.

2.2. Static Detection System Design. The static detection system designed for detecting the metal debris is shown in Figure 2. The system is consisting of probe, metal debris, sensor, impedance analyzer, computer, and 3 axes platform. The metal debris is fixed on the tip of the nonmetal probe. Then the probe is installed on the 3 axes platform. The location of the metal debris relative to the microcoil can be controlled by adjusting the 3 axes platform.

The microinductance coil with 14 turns is fabricated with PCB fabricate process. The width of coil line is 100 μm, and the space between lines is 100 μm also. The photograph of the coil is shown in Figure 3.

The metal debris is fixed on the tip of the nonmetal probe. Then the probe is installed on the 3 axes platform.

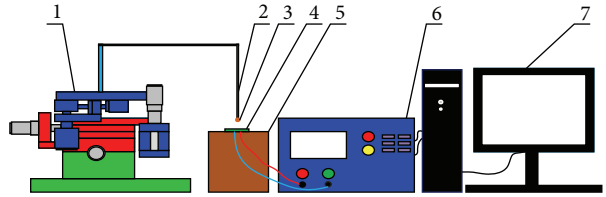


Figure 2: The schematic drawing of static detection system. 1 is the 3 axes platform, 2 is the probe, 3 is the metal debris, 4 is the sensor, 5 is the lug boss, 6 is the impedance analyzer, and 7 is the computer.

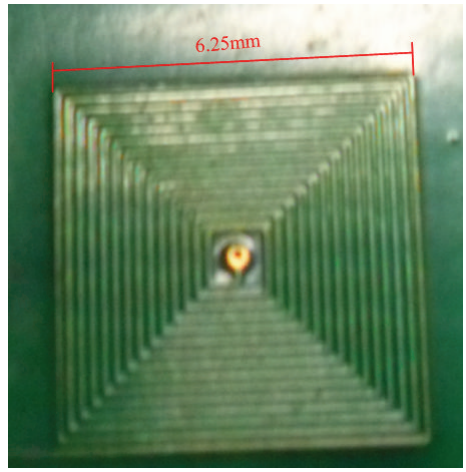


Figure 3: The photograph of microinductance coil.

The location of the metal debris relative to the microcoil can be controlled by adjusting the 3 axes platform.

The metal debris is adjusted on the top of the microcoil, and the distance between the debris and the coil is 200 μm while detecting the inductance of the microcoil. The inductance of coil will be detected while the distance increases by adjusting the 3 axes platform. The detection current frequency is 1 MHz while measuring the inductance of the microcoil.

3. Finite Element Simulation

3.1. Calculation Theory of Eddy Current Field. Alternating magnetic field will appear while alternating current passing through the microcoil. Then the alternating magnetic field will generate an eddy current in the metal debris. The calculation of the eddy current field is to solve the Maxwell equations in the given boundary conditions. The differential form of the Maxwell equations can be expressed as (3) for general time-varying field:

$$\begin{aligned} \nabla \times H &= J + \frac{\partial D}{\partial t}, \\ \nabla \times E &= -\frac{\partial B}{\partial t}, \\ \nabla \cdot B &= \rho, \\ \nabla \cdot D &= 0. \end{aligned} \tag{3}$$

Here H is the magnetic field intensity, J is the current density vector, D is the electric field vector flux density, E is the electric field intensity, B is the magnetic induction, and ρ is the volume density of charge.

The homogeneous wave equations satisfied the needs of 3D-eddy current field are shown in (4):

$$\left\{ \begin{array}{c} \nabla \times \left(\dfrac{1}{\sigma + j\omega\varepsilon} \nabla \times H \right) + j\omega\mu H = 0 \\ \nabla \cdot (\mu \nabla \varphi) = 0 \end{array} \right\}. \tag{4}$$

Here ε is the dielectric constant of medium, μ is the differential permeability, σ is the conductivity of medium, and j is the unit of imaginary.

The magnetic field intensity H is described as (5) in the simulation of Maxwell 3D-eddy current field:

$$H = H_p + \nabla_\varphi + T. \tag{5}$$

Here T is the potential electric vector, φ is the potential electric scalar, and H_p is the magnetic field intensity at the edge.

The value of T is zero while the area is insulated. Only φ and H_p should be calculated. T still should be calculated while the area is conductive.

The boundary conditions and solving precision are set after the model is built. Then the software starts to calculate, generate the initial split of the model and analyze the error. If the result satisfies the terminal conditions, it will be output the result, or else, the mesh will be refined and the field will be recalculated. The computational flow diagram is shown in Figure 4.

3.2. Model Establishment and Parameters Selection. The software of Maxwell is adopted to build and analyze the metal debris detection system. The simulation model is built with the software of Maxwell as shown in Figure 5. The coil is set on the PCB board. The thickness of the PCB board is 0.4 mm. The width of the coil line is 100 μm with 100 μm line space and the thickness is 35 μm. The debris is over the sensor. The debris is gathered in the oil and can be calculated as a metal ball [13]. So the shape of the debris is set as a ball whoes

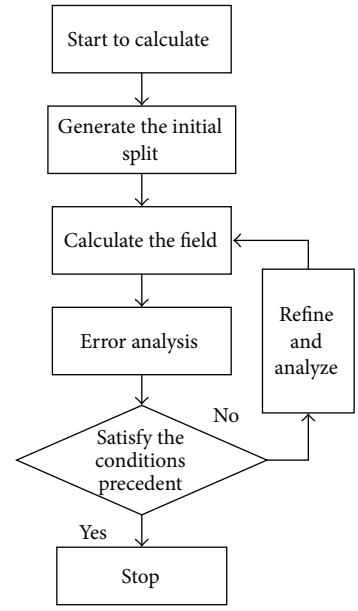


Figure 4: The computational flow diagram.

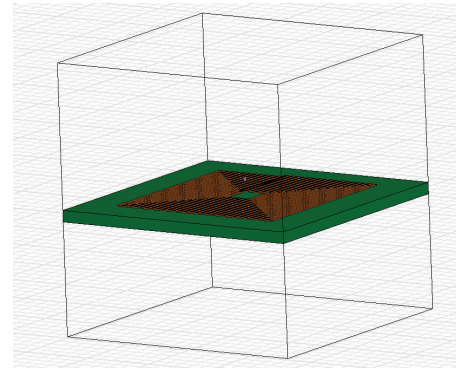


Figure 5: The simulation model of the system.

diameter is 150 μm. The coil is covered with medium and its size is 9000 μm×9000 μm×9350 μm (length×width×height). The whole boundary condition is set as radiation boundary.

The material of the coil is set as copper, the medium is set as air, and the metal debris is set as iron or copper. The magnetic permeability and the electrical conductivity of the material are set as isotropic. The specific parameters of material in the detection model are shown in Table 1.

3.3. Definition the Specification of Solution. The inspirator alternating voltage is 0.5 V and the resistance of the microcoil is 1.088 Ω in the experimental measure of the inductance of the microcoil. So, the peak value of current is set as 0.65 A and the solving frequency is set at 1 MHz. The auto adaptive analysis is selected and the number of the iterations is set as 20 times. The percentage of new mesh will be added after each adaptive loop is set as 30%. The shape of mesh is tetrahedron and the number of mesh is set as 50000. The model is solved and the mesh is refined according to the parameters.

Table 1: Parameters of material adopted in the model.

	Relative permittivity	Relative permeability	Bulk conductivity	Dielectric loss tangent	Magnetic loss tangent
PCB board	4	1	0	0	0
Coil	1	0.999991	58000000	0	0
Debris (Cu)	1	0.999991	58000000	0	0
Debris (Fe)	1	4000	10300000	0	0
Air	1	1	0	0	0

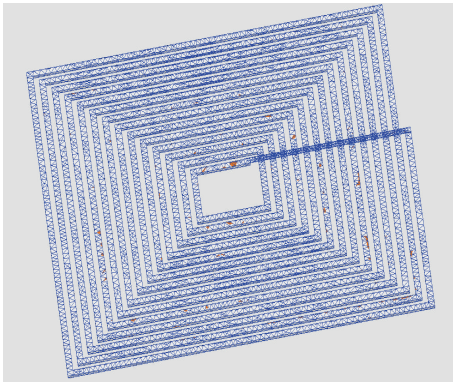


Figure 6: The mesh of the microplane coil.

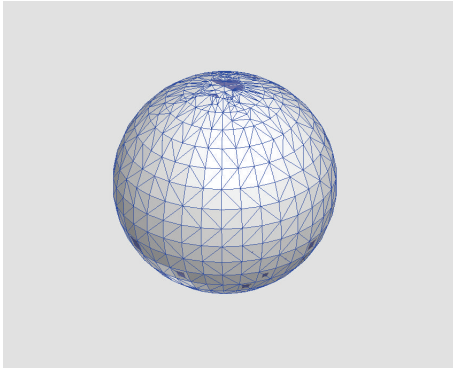


Figure 7: The mesh of metal debris.

The mesh result of the microplane coil is shown in Figure 6 and the mesh result of metal debris is shown in Figure 7.

4. Comparison of Simulation and Experimental

4.1. The Influence on the Magnetic Field. The magnetic field of microcoil is simulated with the model being set up in the software of Maxwell. The debris of iron is set as 150 μm in diameter. The distance between the coil and the iron debris is 200 μm. The debris of copper is set as 150 μm in diameter. The distance between the coil and the copper debris is 50 μm. The magnetic field of microcoil without debris is shown in Figure 8. The magnetic field intensity of microcoil with iron debris is shown in Figure 9. The interior magnetic field of the iron debris is reduced according to Figure 9. The direction of magnetic field generated by the eddy current in the iron

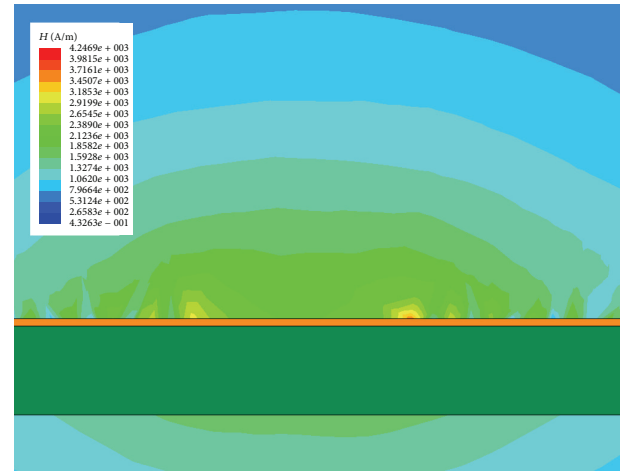


Figure 8: The magnetic field with no debris.

debris is opposite to the magnetic field of the microcoil. So, the total magnetic field is decreased in and near the iron debris. The magnetic induction of the microcoil is shown in Figure 10. However, the inner magnetic induction of the iron debris increases from 3×10^{-3} T to 7×10^{-3} T. Figure 11 is the enlarged drawing of the magnetic induction of the iron debris. The magnetic induction in the debris is not uniform because of the education current of the microcoil. The education current is not helix as the coil. The part of the debris far from the coil has a larger magnetic induction than the side near the coil. The magnetic induction field is also asymmetric due to the asymmetry structure of the coil.

The magnetic inductance field of microcoil with copper debris is shown in Figure 12. The magnetic induction field around and in the debris becomes lower by comparing with no debris. The magnetic inductance near the copper debris decreases from 3×10^{-3} T to 3×10^{-4} T. This is because of the direction that occurred by the eddy current is opposite to the magnetic field of coil. The magnetic field of microcoil with copper debris is shown in Figure 13. The value of magnetic field (H) is similar to magnetic induction field (B) because the relative permeability of copper is approximate to 1.

4.2. The Influence on the Inductance of the Metal Debris. The influence on the inductance of the microdetection coil is simulated with the software of Maxwell. The simulation result of the inductance is 736.783 nH while there is no debris near the coil. The experiment testing value is 705.48 nH while there is no debris near the coil. Copper and iron debris are

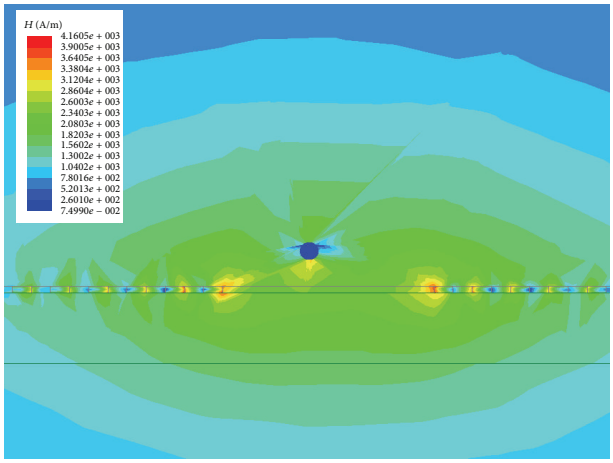


Figure 9: The magnetic field with iron debris.

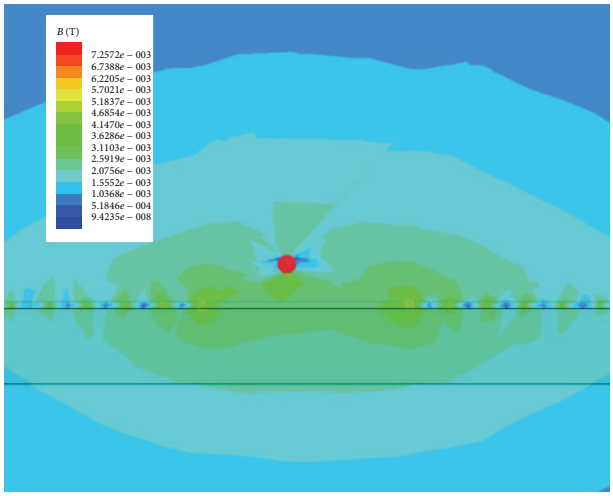


Figure 10: The magnetic inductance with iron debris.

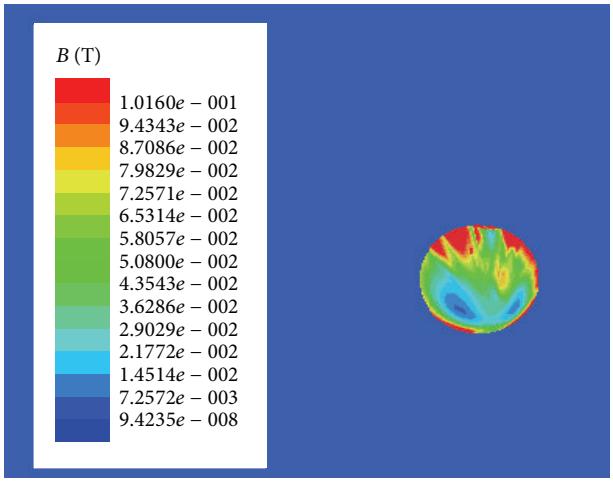


Figure 11: The magnetic inductance of iron debris.

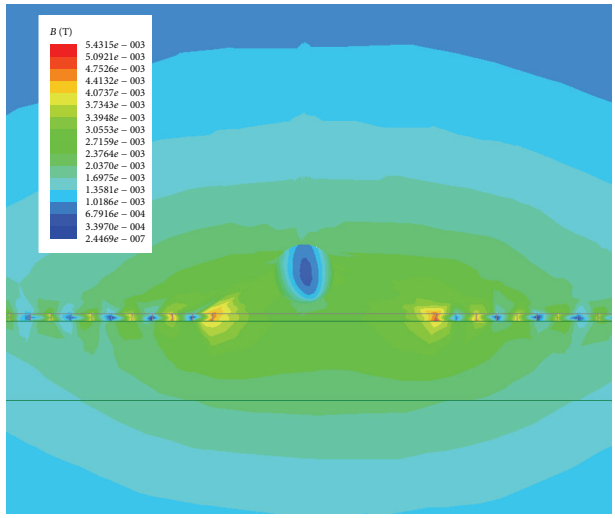


Figure 12: The magnetic inductance field with copper debris.

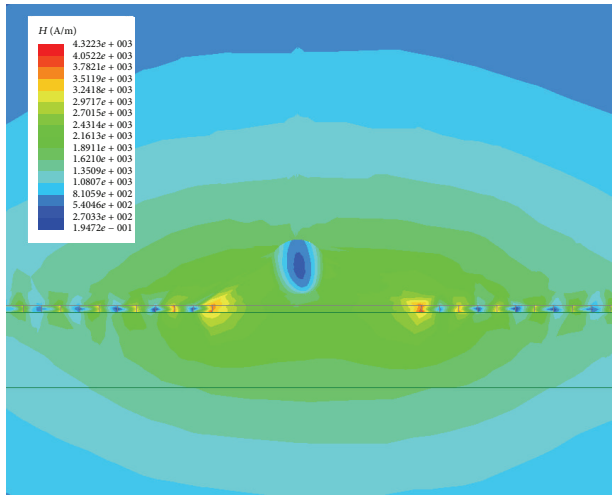


Figure 13: The magnetic field with copper debris.

fabricated for the experimental testing. The micrograph of copper debris is shown in Figure 14, and the micrograph of iron debris is shown in Figure 15. All the debris is about 150 μm in diameter. The experiment is carried on with debris and the static detection system.

The simulation inductance and experimental inductance of microcoil with iron debris are shown in Figure 16. The inductance of the experiment is 706.48 nH and the simulation value is about 737.6615 nH, while the distance between the coil and debris is 200 μm. The inductance is simulated and measured while every 200 μm increased in distance between the microcoil and the debris. Both the simulation and experiment inductance decrease while the distance increases. The relation between the change rate of inductance ($\Delta L/L$) and the distance is shown in Figure 17. The change rate of inductance also decreases while the distance increases. The change rate of inductance in simulation and experimental is consistent.

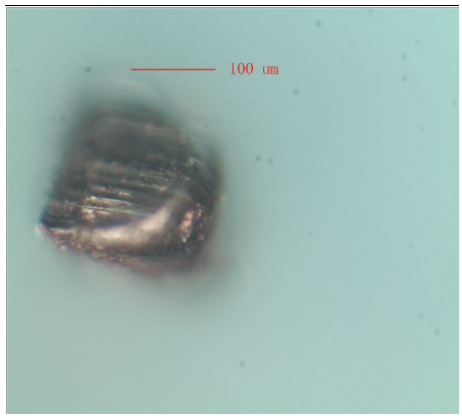


Figure 14: The micrograph of copper debris.

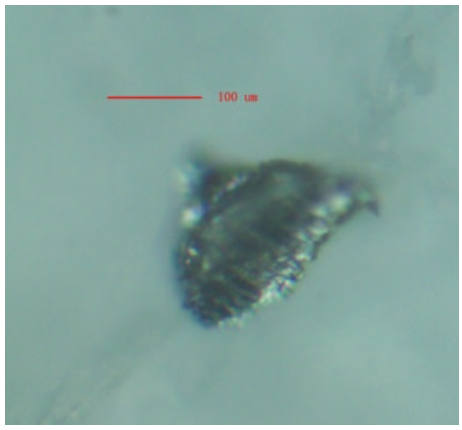


Figure 15: The micrograph of iron debris.

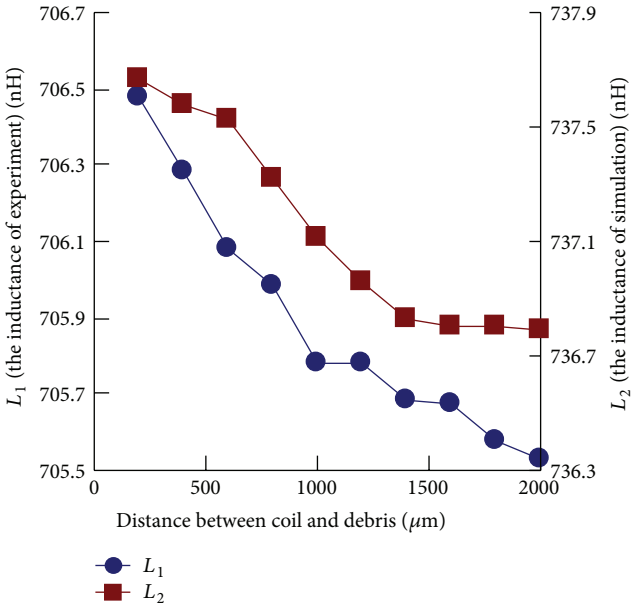


Figure 16: The change of inductance caused by iron debris.

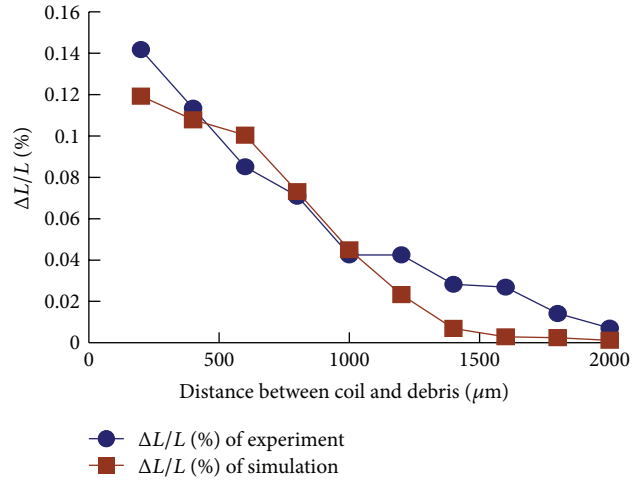


Figure 17: The change rate of inductance caused by iron debris.

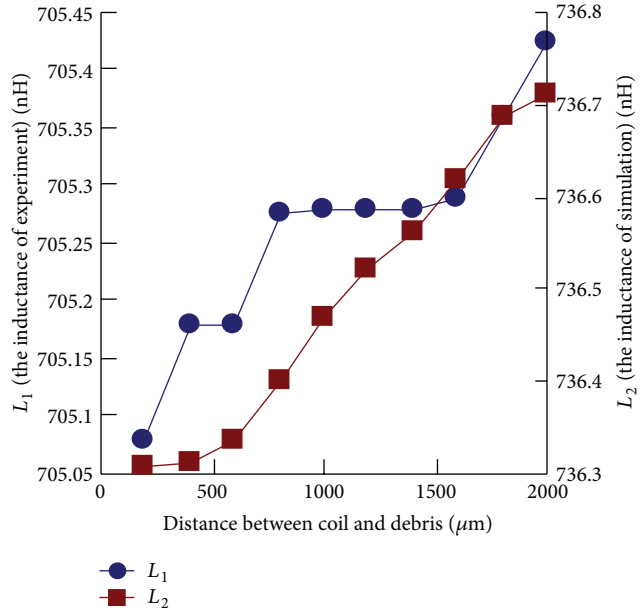


Figure 18: The change of inductance caused by copper debris.

The simulation inductance and experimental inductance of microcoil with copper debris are shown in Figure 18. The inductance of the experiment is 705.08 nH and the simulation value is about 736.308 nH while the distance between the coil and copper debris is 200 μm. Both the simulation and experiment inductance decrease while the distance increases. The relation between the change rate of inductance value ($\Delta L/L$) and the distance is shown in Figure 19. The change rate of inductance in simulation and experimental tests is consistent. The change rate of inductance caused by copper debris is smaller than that caused by iron debris when debris is at the same size.

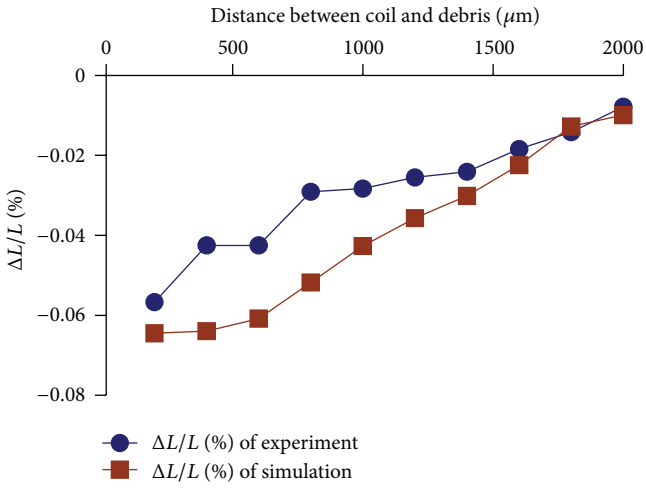


FIGURE 19: The change rate of inductance caused by copper debris.

5. Conclusion and Discussion

The coil of the microinductive sensor is made of copper and its size is very small. The education wire of the sensor and the copper in the sensor will affect the inductance during experimental detection of the metal debris. The excess wires will increase the inductance value of the sensor. Several methods are adopted in order to decrease the measuring errors. The inductance of sensor connected with wires and the inductance of wires are measured, respectively. Then inductance of sensor is obtained with the total inductance subtract the inductance of the wire.

The accurate size of the debris is difficult to be ensured in the production process. Therefore, the debris is observed and selected repeatedly with microscope to ensure that the size is near 150 μm in diameter. The test is carried on in the cleaning room in order to reduce the impact of environment.

The error also comes from the following conditions. The magnetic loss is not considered in the simulation. The material property in the simulation and the actually measuring material property are not entirely consistent. Besides, there is a class of errors coming from the division of the grid and the setting of the error rate in simulation. These may cause the difference between the simulation inductance and the experimental inductance of the microcoil in the detection of the metal debris.

The debris detection system is simulated and analyzed with the software of Maxwell 14 in this paper. The magnetic induction intensity and the magnetic density of the detection system are simulated. This simulation information indicated that the nonferromagnetic debris reduces the inductance of microplane inductance sensor and ferromagnetic debris increases the inductance of microplane sensor. The detection of metal debris with microplane sensor is feasibly proved by the simulation and the experimental test in this paper. This paper provides a model for detecting the debris with a plane eddy current sensor and a case for the 3D simulation of the eddy current. This work may have some significance for improving the efficiency of the plane eddy current sensor.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgments

Project of the National Natural Science Foundation of China no. 51105011 supported this research. This work was supported by the Specialized Research Fund for the Doctoral Program of Higher Education (no. 20101103120002). Project of Beijing Board of education no. KM201210005015 supported this research. This work is done with some help of Beijing key laboratory of Advanced Manufacturing Technology. Thanks are due to Professor Jingpin Jiao for the help in the experimental test.

References

[1] P. Kallappa, C. Byington, and B. Donovan, "Fault detection and lubricant health monitoring for slow bearings in extreme operating conditions," in *Proceedings of the ASME Conference*, pp. 885–886, Washington, DC, USA, September 2005.

[2] J. Han Sen and R. B. Peoples, "Using Eddy current testing to solve industrial problems," *Materials Evaluation*, vol. 64, no. 5, pp. 543–546, 2009.

[3] S. Y. Feng, Y. Zhou, Z. M. Zhou, and Y. Xiang, "The research on properties of micro plane inductance coil which has core," *Electronic Components and Materials*, vol. 18, no. 2, 2009.

[4] J. H. Wang, W. M. Lin, and C. Hong, "The simulation analysis of rectangular spiral integrated inductor," *Jiangsu Electrical Apparatus*, no. 9, 2010.

[5] F. Gao, W. J. Chen, G. F. Li, and K. F. Qu, "The method to calculate multi-conductor partial induction based on the thought of complete collection," in *Proceedings of the CSEE*, vol. 31, December 2011.

[6] D. J. Sadler, S. Gupta, and C. H. Ahn, "Transactions magnetics," *IEEE Transactions on Magnetics*, vol. 37, no. 4, 2001.

[7] H. Fan, Y. Zhang, G. Ren, and F. Chen, "Study on oil detection technology based on inductive wear debris sensor," in *Proceedings of the 9th International Conference on Electronic Measurement and Instruments (ICEMI '09)*, pp. 2810–2813, August 2009.

[8] L. Du, J. Zhe, J. Carletta, R. Veillette, and F. Choy, "Real-time monitoring of wear debris in lubrication oil using a microfluidic inductive Coulter counting device," *Microfluidics and Nanofluidics*, vol. 9, no. 6, pp. 1241–1245, 2010.

[9] L. Du and J. Zhe, "On-line wear debris detection in lubricating oil for condition based health monitoring of rotary machinery," *Recent Patents on Electrical Engineering*, vol. 4, no. 1, pp. 1–9, 2011.

[10] L. Du and J. Zhe, "A high throughput inductive pulse sensor for online oil debris monitoring," *Tribology International*, vol. 44, no. 2, pp. 175–179, 2011.

[11] M. S. Schneider, J. E. Borkow, I. T. Cruz, R. D. Marangoni, J. Shaffer, and D. Grove, "The tensiometric properties of expanded guinea pig skin," *Plastic and Reconstructive Surgery*, vol. 81, no. 3, pp. 398–403, 1988.

[12] W. Tang, L. Zhang, and X. B. Ge, "Finite element simulation of welding crack defect based on Eddy current," *Instrument Technique and Sensor*, vol. 12, 2011.

[13] Z. Peng, N. J. Kessissoglou, and M. Cox, "A study of the effect of contaminant particles in lubricants using wear debris and vibration condition monitoring techniques," *Wear*, vol. 258, no. 11-12, pp. 1651–1662, 2005.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 825872, 6 pages
http://dx.doi.org/10.1155/2013/825872

Research Article

Development of a Smart Residential Fire Protection System

Juhwan Oh,[1,2] Zhongwei Jiang,[1] and Henry Panganiban[3]

[1] *Department of Mechanical Engineering, Yamaguchi University, 1677-1 Yoshida, Yamaguchi Prefecture, Yamaguchi-shi 753-8511, Japan*
[2] *Masteco Industry CO., Ltd., Room 1009, Byeoksan Digital Valley 5, 60-73 Gasan-dong, Geumcheon-gu, Seoul 153-801, Republic of Korea*
[3] *School of Mechanical and Automotive Engineering, Kunsan National University, 558 Daehakro, Gunsan, Jeonbuk 573-701, Republic of Korea*

Correspondence should be addressed to Zhongwei Jiang; jiang@yamaguchi-u.ac.jp

Received 22 March 2013; Revised 10 June 2013; Accepted 23 June 2013

Academic Editor: Yong Tao

Embedded system is applied for the development of smart residential fire detection and extinguishing system. Wireless communication capability is integrated into various fire sensors and alarm devices. The system activates the fire alarm to warn occupants, executes emergency and rescue calls to remote residents and fire-fighting facility in an intelligent way. The effective location of extra-sprinklers within the space of interest for the fire extinguishing system is also investigated. Actual fire test suggests that the developed wireless system for the smart residential fire protection system is reliable in terms of sensors and their communication linkage.

1. Introduction

Every year residential fires occur practically around the world. For example in USA, recent estimates showed an average yearly of 371,700 residential structures fires causing 2,590 civilian deaths, 12,910 civilian injuries, and $7.2 billion in direct property damage. Majority (71%) of the reported residential fires occurred in one- or two-family homes with the remaining occurred in apartments or other multifamily housing [1]. In the UK, 43,500 residential (dwelling) fires were reported for the year 2011-2012. These fires caused 287 fatalities and 11,300 nonfatal casualties [2]. The damages to property, injuries, and loss of lives caused by residential fires have great impact on the economy and society. This impact has continuously encouraged the promotion and implementation of stricter fire safety requirements particularly in industrialized and advanced countries. With these facts, it is imperative that there is a persistent need for the development of technology for the protection of residential buildings against fire.

In an event of residential fire, smoke alarm can save lives and reduce the loss to properties by providing early warning alarms to residents. In many cases, smoke alarms are combined with other detection technologies (such as gas and heat sensors) to ensure an efficient and reliable detection of the early indicators of fire occurrence [3–6]. For extra protection against fire, however, in addition to smoke alarms, installation of fire sprinklers is recommended. It is reported that the risk of dying and property loss decreases by about 80% and 71%, respectively, when sprinklers are installed [7]. A wise approach, therefore, is to combine smoke alarms (and other sensors) with fire sprinkler system to obtain a reliable and more efficient fire protection system. Recently, the addition of fire sprinkler has become part of the building requirements for residential structures [8].

Recently, wireless system (ZigBee) has also received attention for its application in the fire early detection system [9]. This system is becoming more attractive to housing which accommodates elderly couple or elderly person living alone because it is perceived to be more secured. The fire alarm system using wireless technology is seen to be necessary in modern society as part of the fire-fighting equipment. In addition, inspired by the era of communication and data sharing, the wireless home network applications are now freely available. While the current communication technology within the fire-fighting industry is mostly still based on wired system, there is a growing need for the application of wireless communication between fire-fighting system components with the capability of home network

integration. ZigBee-equipped wireless smoke sensors along with IT technology can activate alarms if a fire occurs. In this study, the ubiquitous technology is implemented for the purpose of ensuring fire safety in residential buildings through wireless fire detection and extinguishing system. This system also aimed at reducing installation cost due to space restriction and promoting ease and practicality of installation in existing residential buildings. Unlike wired interconnection which may only be practical for use in new construction, especially if the wires are laid out without cutting walls and ceilings (or floors in multistory residence), the wireless interconnected sensors can be conveniently retrofitted in buildings without costly wire installations. In effect, the building interior design is not compromised and total installation cost can be significantly reduced.

The main ideas of the proposed fire detection and extinguishing system were as follows: Firstly, in housing condition where it is not easy or impractical to install conventional fire detection equipment, a smart fire sensor network using wireless communication system that activates and sends warning alarm within the building, to the building residents or owner, and the fire department was developed. Secondly, to implement an improved compact water-based fire-extinguishing sprinkler system with higher discharge so that the fire can be effectively extinguished at the early stage. The compactness of the extinguishing system is believed to cut the total installation cost. The two systems were integrated for the development of practical wireless fire detection and extinguishing system for residential building applications. Experimental equipment for water supply with specific type of sprinklers and wireless communication were developed and tested to verify the system performance.

2. System Workflow

In residential structures, a potential fire can be indicated by the presence of smoke, heat, gas, or their combination. The smart fire protection system presented in this paper is initiated upon the detection of these early indicators of fire. Generally, a warning signal or alarm is produced when the concentration of detected smoke, carbon monoxide (CO), liquefied natural gas (LNG), or liquefied petroleum gas (LPG) has met the criteria for alarm activation, that is, above threshold values. These sensors are interfaced with the control system so an automatic fire notification system is executed at the instant the fire alarm is activated. Telephone networking is included in the control system thereby automatic notification call or message is sent to the fire department and residence owners who are off the premises. In an unlikely event of fire, the sprinkler system is activated allowing water to flow out of the sprinkler heads to extinguish the fire. Since the residential sprinklers are temperature-activated, only the sprinkler head closest to the fire source is set off, thus, reducing damage to properties. Water pressure information in the sprinkler line is sent into the control system through the standalone type alarm valve. The water pressure required to extinguish the ongoing fire is supplied by a water pump that is also interfaced with the control system. When fire is totally suppressed, the water supply is automatically shut off. In case that LNG or LPG leakage is detected which is attributed by the gas concentration above the preset limits of the sensor, the gas line is automatically shut off and the gas bureau and residence owner are instantaneously notified. Essentially, the system operation is illustrated in Figure 1.

To minimize false fire alarm and rescue calls, the smart system has particularly utilized an integrated CO + Smoke sensor which follows an algorithm for identifying actual fire based on the combination of detected smoke and CO concentration. This is mainly based on the fact that CO is a product of an incomplete combustion that is characterized by the presence of smoke. In other words, the amount of CO present in the detected smoke is a determining factor for fire detection, fire extinguishing system activation, and subsequent notification calls. Various other algorithms for similar kind of sensor integration for fire detection are available in the literature (see [3, 6] among others). In the developed residential fire detection and extinguishing system presented herein the corresponding sensor triggers the alarm when threshold value is reached (as depicted in Figure 1). These threshold values are preset by the manufacturer. Additionally, the CO + Smoke sensors installed in the whole area can be wirelessly interconnected with each other, so that if one detects smoke or confirms fire, the alarm will sound on all sensors in the network. This setup improves the ability to get the residents alerted either they are on closed-door bedrooms or living two floors away from the location. Wireless linkage and networking of the CO + Smoke, LNG/LPG, and water pressure sensors with the control system is implemented using ZigBee technology as depicted in Figure 2. By having the sensors wirelessly interconnected, the technology allows the sensors to be retrofitted in a residential building without costly wire installations.

Conventional sprinkler system often gets its water supply from an overhead water reservoir or through connection to commercial water line which may need pump to maintain the required water pressure. The sprinkler system for this proposed smart fire protection system has implemented a compact cabinet-type design. That is except for the sprinkler heads and the corresponding water pipelines, all components of the sprinkler system such as water tank, pumps, pressure gauges, and valves, and even the control panel are housed in a steel cabinet. This feature has made the system highly mobile, and can be conveniently installed in residential buildings and transferred from one property to another. The fire extinguishing sprinkler system is depicted in Figure 3.

The smartness of the developed system can be described by its ability to discriminate or make decisions when to trigger the necessary alarms, execute notification calls, or activate and regulate the fire extinguishing system automatically. Furthermore, it is capable of monitoring and storing information of all events in the system.

3. System Components

3.1. CO + Smoke Sensor. Existing standalone alarm-type fire sensor is basically capable of fire detection through smoke

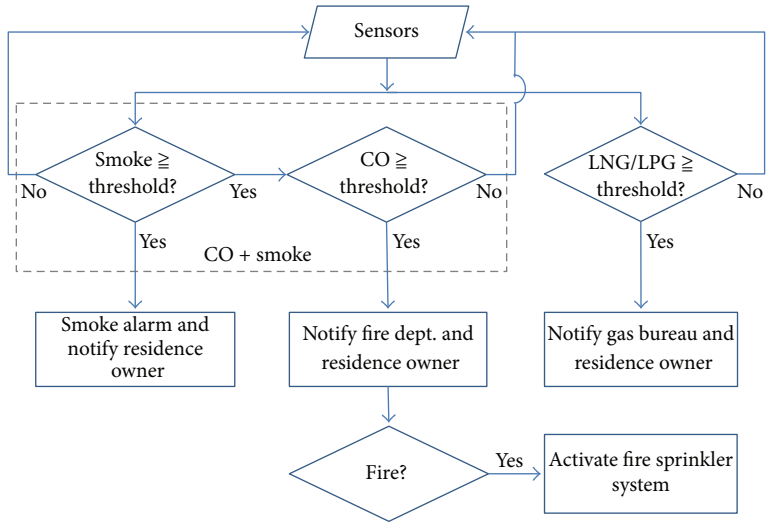


Figure 1: Fire protection system operation schematic.

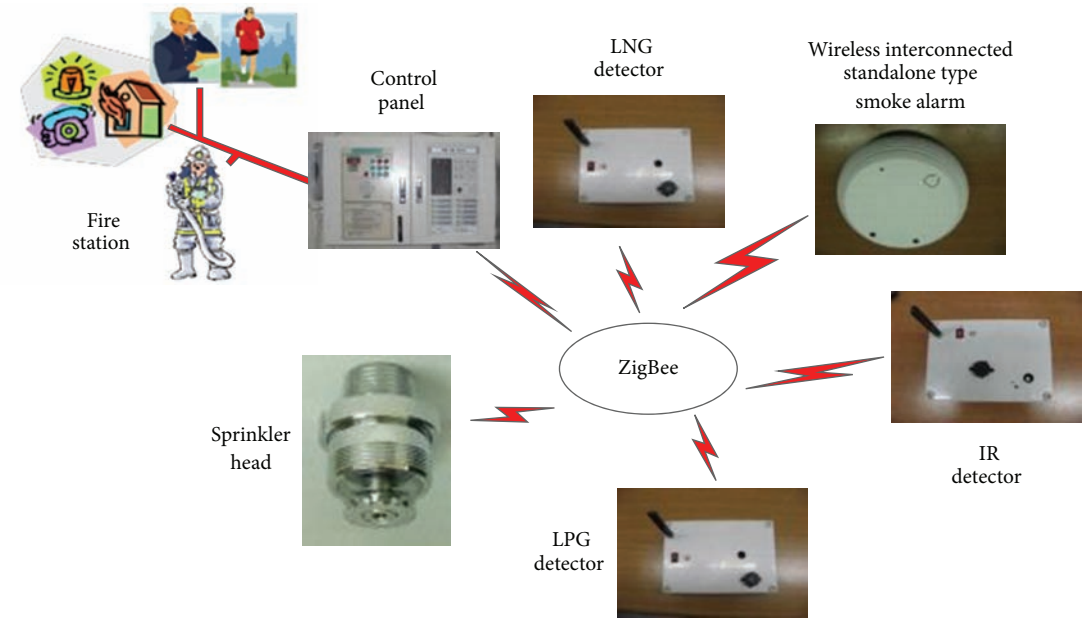


Figure 2: Wireless linkage system.

sensors with local alarm. In this study, wireless CO + Smoke sensor used for the proposed fire protection system enables wireless interconnection with other sensors which has added safety feature as has been described in the preceding section. It basically consists of two sensors, CO and infrared (IR) smoke sensors packed into a single unit. Using the compact single-unit CO + Smoke sensor is found to be less expensive than having two sensors installed separately. Wireless buzzer is added to activate sound alarm once detection is confirmed. Table 1 shows the specification of the CO + smoke sensor.

3.2. LNG/LPG IR Sensor. LNG or LPG leakage in the residential buildings is potentially hazardous. For instance, natural gas leaks can cause immediate death if high concentration is inhaled. These gases are highly flammable. The developed system includes LNG and LPG IR sensors. Table 2 shows the specification of the LNG and LPG sensors.

Table 1: Specification of the CO + Smoke sensor.

Sensor	CO	IR Smoke
Threshold	0.055% (550 ppm)	15% obs./m
Alarm	Buzzer (1 mW, 54 mA)	
Transmission	ZigBee (2.4 GHz, 9600 bps)	
Size (mm)	50 H × 120 D	

3.3. Wireless I/O and TCP/IP Interface. Wireless sensors are linked to the control panel through wireless I/O module.

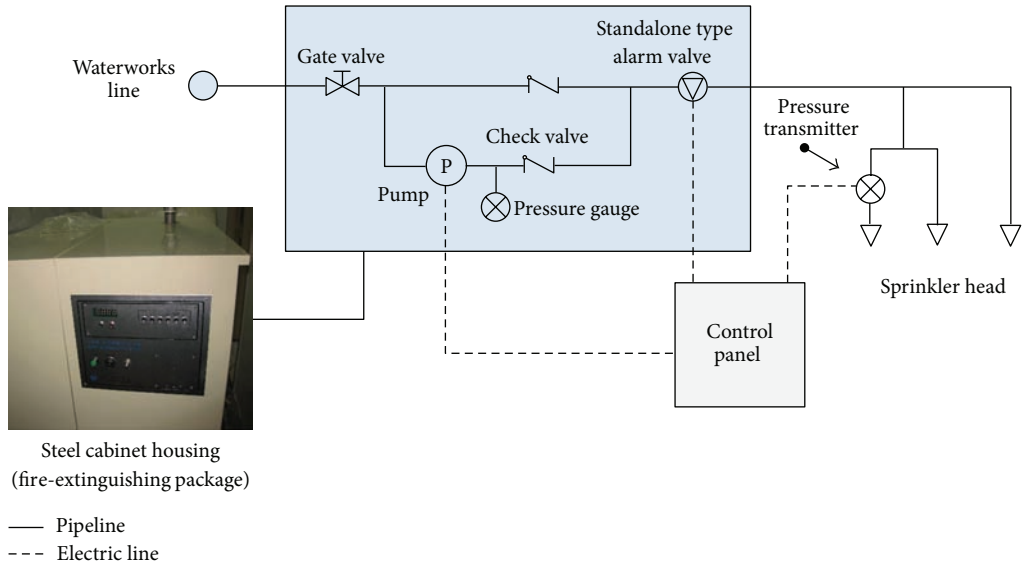


Figure 3: Sprinkler system connection with the control system.

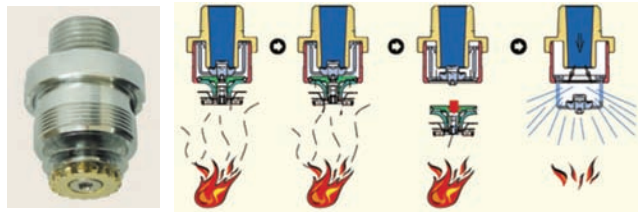


Figure 4: Flush-type sprinklers (leftmost) and its activation mechanism during fire.

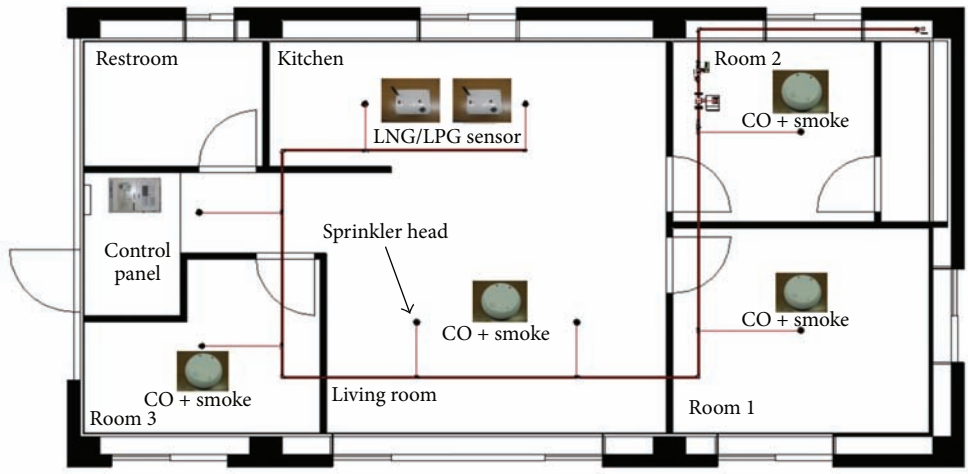


Figure 5: Test room configuration for wireless linkage and sprinkler system test.

ZigBee wireless technology is used to receive wireless signal from each sensor and initiate the alarm system in case of fire. The test result obtained with the communication program indicates that the wireless module can cover area within 7-meter radius. The notification system is connected with the control panel through serial ports. Through telephone line, the system automatically calls the fire department for notification and fire extinguishing assistance if needed. Subsequently, the residence owners are also automatically notified. Emergency notification system includes voice transmission, manual call, and communication recorder. Table 3 shows the wireless I/O and TCP/IP module specifications.

Figure 6: Fire ignition and suppression sequence during the actual fire test.

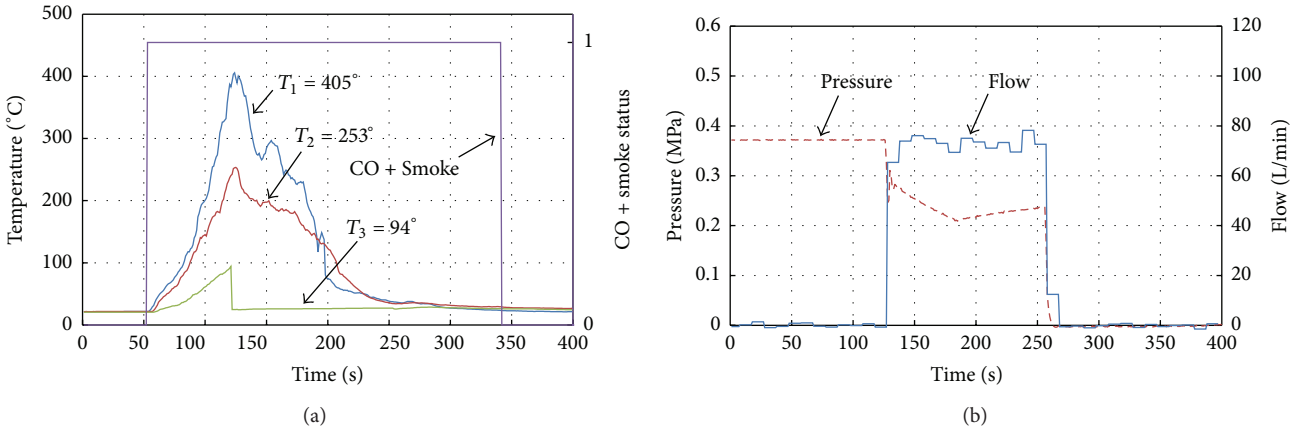


Figure 7: (a) Temperatures at various test points during fire suppression with the corresponding CO + smoke sensor status monitor (1-active; 0-standby) and (b) water pressure and flow rate data at fire suppression.

Table 2: Specification of the LNG and LPG sensor.

Sensor	LNG	LPG
Threshold	1.25%	0.45%
Current	120 mA	
Transmission	ZigBee (2.4 GHz, 9600 bps)	
Size (mm)	180 W × 130 H × 35 D	

Table 3: Wireless I/O and TCP/IP module specifications.

Module	I/O	TCP/IP
Current	100 mA	200 mA
Transmission	Zigbee	Serial
Size	53 W × 40 D	65 W × 18 H × 42 D

The CO + Smoke sensor with the wireless module was tested at a house fire environment. It was still operational and works stable when the temperature was over 400°C. Furthermore, the wireless model has installed the reciprocal checking function and they will communicate with each other periodically. If there is no response from the neighbor sensor, a notification will be sent to the residents or owner. If one sensor module is in alarm status the other alarms will be activated also.

3.4. Sprinkler System. Existing sprinkler system is used for the proposed residential fire protection system. For the purpose of testing, flush type sprinklers are used. Figure 4 shows example of the flush type sprinkler head and its activation mechanism (fuse metal melts at known temperature, set off the heat collector and discharges pressurized water).

4. System Testing

4.1. Wireless Communication Linkage Test. Figure 5 shows the location configuration of the sensors and control panel installed for wireless communication linkage test. A monitoring program was developed for testing purpose. To test the functionality of the devices described earlier, smoke and gas were introduced into each area where corresponding sensors were installed. Once the triggering substance is detected, wireless signal was received and the fire alarm was activated. The test has been shown to confirm the wireless communication capability of the system. In the fire test with a single-floor three-bedroom residential building, four CO + smoke sensors and one LNG/LPG sensor were found to be the optimal configuration. In such sensor location, configuration wireless communication can be established within 7-meter radius.

4.2. Fire and Wireless Linkage Test. At high-temperature environment, the operational reliability of the sensors and linkage capability are highly critical. Hence, the system linkage test was carried out with fire occurrence scenario. In this study, computer program was developed for the room fire monitoring test. The program also facilitates the connection between the fire detection system and automatic fire notification system. Figure 5 shows the layout of the sprinkler heads in the test room. Wooden materials commonly found in residential buildings were contained in a metallic bin situated in the room as shown in the leftmost of Figure 7. The wooden material (fuel) was manually ignited. As the smoke ascended and spread into the ceiling and reached the CO + Smoke 52 seconds after ignition the fire alarm was triggered. The status monitor reported an active status (active = 1; standby = 0) of the CO + smoke sensor. The fusible material in the nearby sprinkler head melted when ceiling temperature reached 405°C opening the head and allowing water to spray down into the fire source. Figure 6 depicts a sequence of fire buildup, sprinkler head activation, and finally fire suppression.

As can be seen in Figure 7(a), smoke detection status continued to be reported while the fire was ongoing with temperature over 400°C which appears to indicate that the CO + Smoke sensor linkage was operational and stable at such critical condition. Due to water flow in the pipe line, standalone alarm-type valve sent pressure signal to the control panel. The control panel automatically adjusted the pump speed so that the required instantaneous water flow and pressure were supplied into the sprinkler. As the fire temperature gone down which indicated fire suppression, the amount of water being discharged was properly regulated. Test revealed that fire is completely extinguished within 247 sec (4.11 minutes) as shown in Figure 7.

5. Conclusions

In this study, the development of a wireless fire detection and extinguishing system for residential application has been presented. Wireless connectivity and communication capability were added to the basic features of various kinds of fire sensors using ZigBee wireless technology. The developed system allows wireless home networking that can send fire notification to residents and fire department automatically. Through actual fire test, the performance of the integrated wireless fire detection and extinguishing system has been verified even at critical conditionhigh temperature condition. Wireless communication among devices in the fire test room was established within 7-meter radius. The improved compact fire extinguishing system has the ability to suppress fire at the early onset within less than 5 minutes. The result of this study can be used for the development of smart, effective, and efficient fire safety and fire-fighting system to address the related demands of a modern society.

References

[1] M. Athens, *Home Structures Fires*, National Fire Protection Association, 2012.

[2] Fire Statistics, in *National Statistics*, Department for Communities and Local Government, London, UK, 2012.

[3] S.-J. Chen, D. C. Hovde, K. A. Peterson, and A. W. Marshall, "Fire detection using smoke and gas sensors," *Fire Safety Journal*, vol. 42, no. 8, pp. 507–515, 2007.

[4] B. Ko, K.-H. Cheong, and J.-Y. Nam, "Early fire detection algorithm based on irregular patterns of flames and hierarchical Bayesian Networks," *Fire Safety Journal*, vol. 45, no. 4, pp. 262–270, 2010.

[5] D. Gutmacher, C. Foelmli, W. Vollenweider, U. Hoefer, and J. Wöllenstein, "Comparison of gas sensor technologies for fire gas detection," *Procedia Engineering*, vol. 25, pp. 1121–1124, 2011.

[6] D. T. Gottuk, M. J. Peatross, R. J. Roby, and C. L. Beyler, "Advanced fire detection using multi-signature alarm algorithms," *Fire Safety Journal*, vol. 37, no. 4, pp. 381–394, 2002.

[7] J. Hall, *U. S. Experience With Sprinklers*, National Fire Protection Association, 2012.

[8] D. Wotapka, "Builders smokin' mad over new sprinkler rules," *The Wall Street Journal*, 2010.

[9] L.-C. Huang, H.-C. Chang, C.-C. Chen, and C.-C. Kuo, "A ZigBee-based monitoring and protection system for building electrical safety," *Energy and Buildings*, vol. 43, no. 6, pp. 1418–1426, 2011.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 783673, 10 pages
http://dx.doi.org/10.1155/2013/783673

Research Article

Optimal Partial Reconfiguration for Permanent Fault Recovery on SRAM-Based FPGAs in Space Mission

Jie Zhang,[1] Yong Guan,[2] and Chunjing Mao[2]

[1] *School of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China*
[2] *School of Information Engineering, Capital Normal University, Beijing 100048, China*

Correspondence should be addressed to Jie Zhang; jzhang@mail.buct.edu.cn

Received 20 April 2013; Accepted 18 May 2013

Academic Editor: Hongxing Wei

In this paper, we present a technique to maximize the lifetime of SRAM-based FPGAs in space mission. We focus on recovering permanent faults induced by SEE (single-events effect). In our technique, we use a fix-sized fault detection module to detect permanent faults and propose a permanent fault recovery mechanism for fault recovery. By using partial reconfiguration, we develop a system lifetime estimation model to find the optimal partition for designing the module-based fault recovering with the maximum system lifetime. We conduct experiments with a set of real applications including SpaceWire, Wavelet, AC97, MEPG-4, 8086, and Ethernet on Xilinx XUP platforms. The experimental results show our technique can effectively improve the lifetime compared with the previous work.

1. Introduction

Since the first prediction of SEE (single-event effect) in microelectronics in 1962 by Wallmark and Marcus [1], researchers and engineers who work on space-based systems have been concerned about the effects of cosmic rays on mission reliability. Radiation environment in aerospace is shown in Figure 1. This concern was realized in 1975 when an anomaly in a spacecraft system was first attributed to the passage of an energetic heavy-ion [2]. Since that time, a long list of single-event effects in semiconductor devices and integrated circuits has been compiled. Some of these effects result in a "transient fault," which causes no permanent damage and can be reset by applying the correct signals to the device. Other effects so-called "permanent fault" can damage the semiconductors permanently. Although these permanent faults can be mitigated by using radio-hardened components, radio-hardened components have a very low performance and cannot meet the requirements in current space exploration mission. Hence, new substitutes with both high reliability and high performance are needed in space mission.

FPGAs (field programmable gate arrays) have many advantages over microprocessors for space-based applications. A variety of projects have demonstrated the benefits by using FPGAs in spacecrafts [3, 4]. For example, the Mars rovers use FPGAs for motor control and landing pyrotechnics [5]; the Australian satellite FedSat uses FPGAs as part of its high-performance computing payload [6, 7]. SRAM-based FPGAs are especially appealing due to their in-site reconfigurability and high performance for signal processing tasks. While SRAM-based FPGAs offer a number of unique benefits for spacecraft electronics, they are also susceptible to SEEs, which induce both transient and permanent faults in SRAM-based FPGAs. While transient faults can be recovered by reconfiguring the same resources, permanent faults require spare resources allocated in the FPGA to replace faulty resources. Therefore, partition and placing of an application with partial reconfiguration can affect the system lifetime.

Several approaches have been proposed to solve transient faults by using partial reconfiguration in SRAM-based FPGAs. A method using read-back and scrubbing is introduced in [8]. Read-back is based on periodically reading the configuration memory of the FPGA to detect if a transient

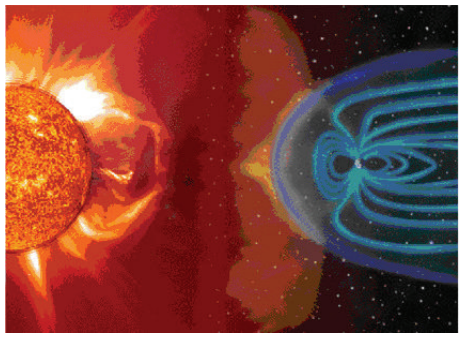


Figure 1: Illustration of radiation environment in aerospace.

fault has occurred. This technique can fix errors by creating the necessary reconfiguration bitstream, through a small bit manipulation approach [9], and comparing the read-back file and the original one. Scrubbing [10] is a method to correct SEUs, in which it performs a periodic reloading (usually with the same frequency of read-back) of the entire configuration bitstream, no matter whether an error has occurred or not. In [11], a duplication-based method is proposed to recover transient faults. The method applies concurrent error detection techniques to monitor the health of the system and to trigger the reconfiguration for the part with faults. In [12], the authors proposed a fault-tolerant reconfigurable architecture based on island-style FPGAs. This architecture contains "autonomous-repair cells" that can repair transient faults autonomously and dynamically. All of the above approaches cannot recover permanent faults.

There have been studies to solve permanent faults. In [13–15], a method based on STARs (self-testing areas) is proposed to test specific offline actions of a device, and the STARs are moved across the FPGA to cover all the off-line sections [13–15]. Although STARs can mitigate permanent fault, they introduce a big time overhead and cannot correct permanent faults in module-based FPGAs. A tile-based solution is presented in [16–18]. In this technique, an FPGA is divided into small partitioned tiles/blocks that have fixed interfaces to other tiles. Diagnosis can locate faulty resources with a granularity smaller than the dimension of a tile, such that faulty resources can be replaced with the spares in the tile. In [16, 18], a hierarchical model is proposed. The model has two hierarchical levels of redundancy: at the lower level, the FPGA is organized with fixed tiles, and each tile includes spare resources; at the higher level, faulty tiles can be replaced with spare tiles.

In [16, 19, 20], a coarse-grained redundancy model is presented, in which spare resources are lumped into tiles or columns. When faults are detected in a column, the whole column is marked as faulty and replaced by a spare column. In [21], the authors presented an efficient on-line failure detection method and integrated it into a reconfigurable system for executing and testing multiple automotive inner cabin functions. These methods also allow a certain degree of failure recovery, and they may provide solution that can heal the system itself from advanced faults. By using the strategy of redundancy, several techniques have been proposed to mitigate faults [22–26]. The triple modular redundancy (TMR) is one of the most commonly used methods, in which three identical functional blocks are utilized for the same functions and their outputs are checked and decided by a majority voter. The TMR introduces too much cost overhead [27]. In all of the above methods, the system lifetime is not considered.

In this paper, we propose a reconfigurable module-based fault recovery mechanism on SRAM-based FPGAs. We divide an FPGA-based application into a number of function modules and add a fix-sized fault detection module for each function module. We present a lifetime model to show the lifetime of an application, and based on this, an optimal partition can be obtained with different sizes of function modules.

To evaluate the effectiveness of our proposed technique, we conduct experiments with six real applications on Xilinx XUP platforms. The experimental results show that our technique can effectively improve the system lifetime compared with the previous work.

The remainder of the paper is organized as follows. In Section 2, we present the background. We introduce the system lifetime model and our optimal partition algorithm in Section 3. In Section 4, we present the experimental results. In Section 5, we conclude the paper.

2. Background

In this section, we present the background for the space radiation and its effects on FPGAs.

Integrated circuits operating outside the earth's atmosphere are exposed to a radiation environment that is much different from the radiation found on earth. Figure 2 shows the distributing of energetic heavy-ion in different altitudes above the earth. It can be observed that the radiation level increases with the altitude.

High-level radiation may influence normal operations of a semiconductor-based device. SEE is caused by protons and heavy ions emitted by the sun (i.e., solar particles), galactic cosmic rays, and particles trapped in the earth's magnetic field [28]. SEE causes both transient and permanent faults on electronic components. Transient faults can be reset by applying the correct signals to electronic devices. Typical transient faults include SEU (single-event upsets), SEL (single-event latchup), and SEFI (single-event function interrupt). Permanent faults cause permanent destruction of electronic devices and cannot be recovered. Typical permanent faults include SEB (single-event burnout), SEGR (single-event gate rupture), and SESB (single-event snap back).

With respect to radiation, SRAM-based FPGAs suffer from the same problems as other semiconductor devices. For transient faults, we can recover it by reconfiguring resources, while permanent faults require spare resources allocated in the FPGA to replace faulty resources after permanent faults are located. In this paper, we focus on permanent faults.

SRAM-based FPGAs are suitable for space mission with high density and reconfigurability compared with radiation-hardened antifuse parts. An FPGA consists of many basic configurable logic blocks (CLBs); CLB is the basic configurable unit, so the size (or the occupied number of

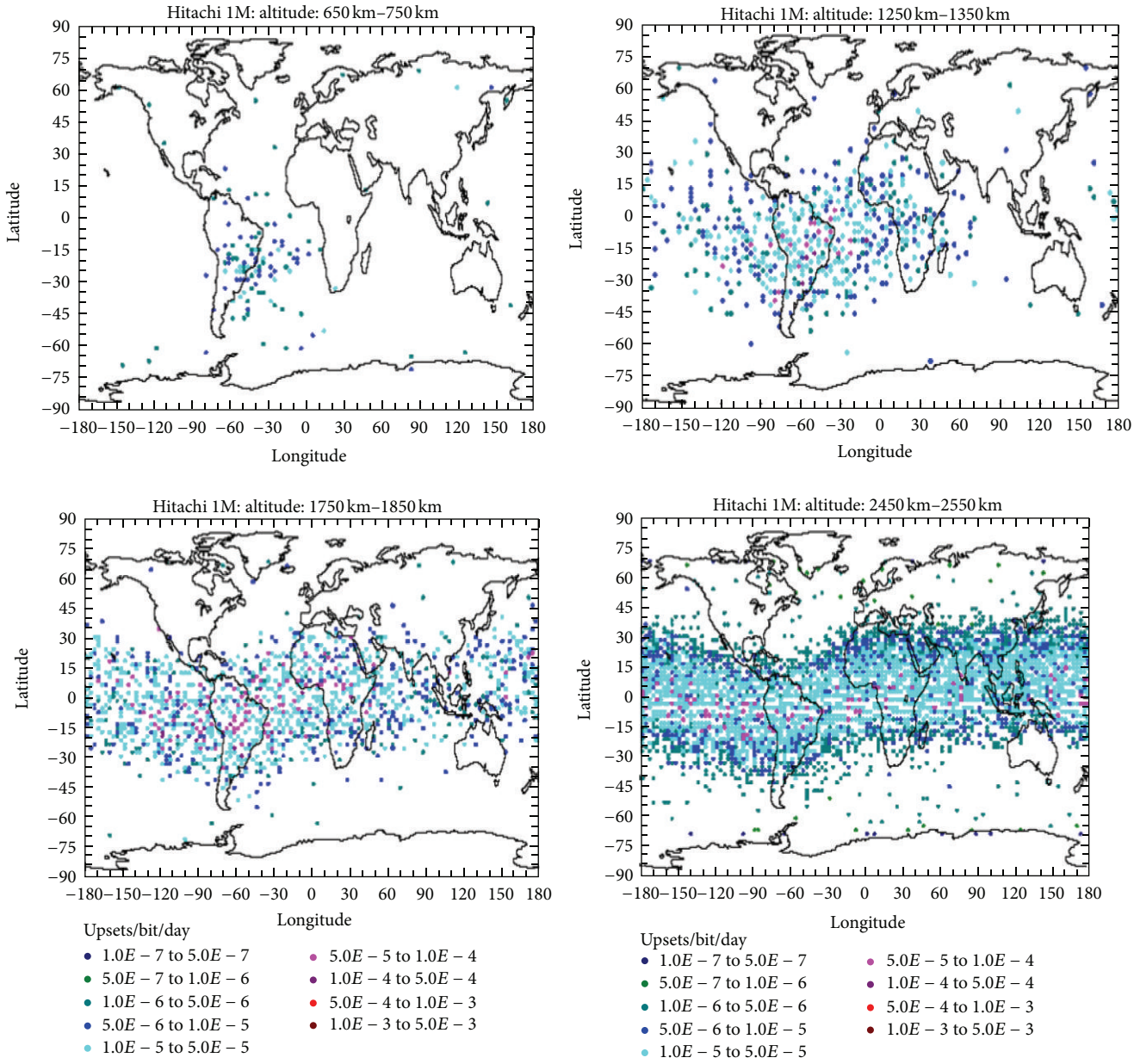


Figure 2: Illustration of a radiation environment in aerospace.

CLB) of each reconfigurable module should be the multiple of the size of one CLB. One SEE-induced permanent fault can only lead one CLB to be damaged. Nonoccupied resources and connections among CLBs are fault-free from the radiation effects. By utilizing a set of fault-free resources and excluding the faulty ones, faulty resources can be replaced by partially reconfiguring the FPGA with an alternative configuration that preserves the same logical functionality.

3. Fault Recovery

In this section, we present our fault recovery mechanism. A given application can be divided into independent function modules, and each module is with the same size. For each function module, a fault detection/location module is added for fault detection and location. Therefore, an actual module with fault tolerance consists of a function module and a fault detection/location module. Without loss of generality, we name actual modules as $m_1, m_2, m_3, \ldots, m_i$.

Following the size of an actual module, we divide the FPGA into a number of cells and sequentially mark cells as $c_1, c_2, c_3, \ldots, c_i$. The size of each cell is the same and equals the size of an actual module. As randomly placing actual modules may lead to routing problem, we first calculate how many cells will be sequentially used. Then we place the application in the FPGA following the shape and the topology

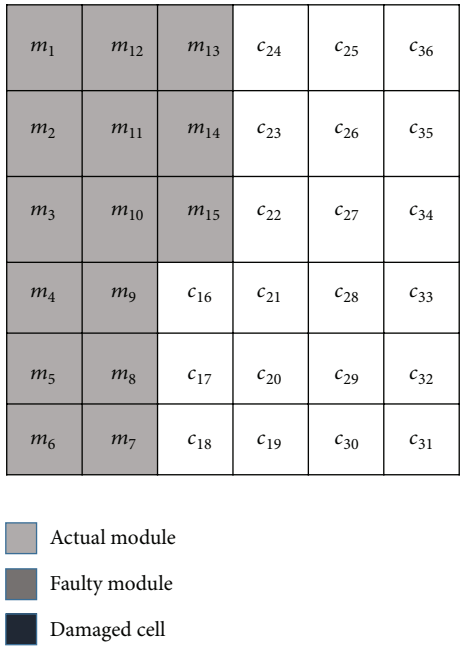


Figure 3: Illustration of cells partition and module placement.

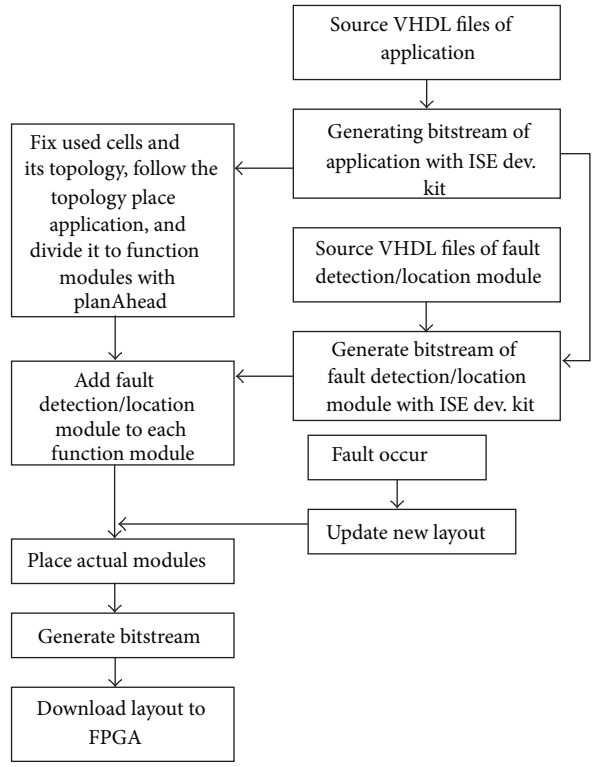


Figure 4: The work flow of our fault recovery mechanism.

of the used cells and divide them into function modules with Xilinx PlanAhead. For each function module, we add a fault detection/location module. With PlanAhead, we place actual modules in sequential cells in an FPGA by matching the SN (serial number) of actual modules and cells, as illustrated in Figure 3.

The work flow of our fault recovery mechanism is shown in Figure 4. Basically, when faults have been detected in one module, we will shift this module and all modules following it to available cells. As illustrated in Figure 5, if we detect that m_9 is faulty, we will shift m_9 and its following models, $m_{10}, \ldots, m_{15}$, to available cells one by one. Then we replace and reconfigure the new layout on the FPGA. This fault recovery mechanism repeatedly works while permanent fault occurs in any module until all available cells have been used up. Due to abundant routing resources in an FPGA, in our experiments, we found that every different layout can work well without routing/interconnection problems in six real applications.

4. Optimal Partition

In this section, we present a technique to find the optimal partition to maximize the system lifetime. We first propose a lifetime model and show how to find the maximal system lifetime and its corresponding optimal partition. Finally, an example is given to illustrate the technique.

The size of a given FPGA is represented as S, the size of an original application system is A, and the size of the fault detection module is f. Suppose we can divide an original application system into n function modules, and the size of a function module is a, where $A = a \times n$. Because of the characteristic of FPGAs, a should be the multiple of integers.

The size of an actual module is $m = a + f$. The size of an application with the fault detection modules is

$$A_f = n \times m = n(a + f). \tag{1}$$

The free size of the FPGA after the initial configuration is

$$R = S - A_f. \tag{2}$$

Because the resources of an FPGA are limited, the reconfiguration/recovery times are also limited. In our technique, the system recoverable times are

$$N = \frac{R}{m} = \frac{S - n(a + f)}{a + f}. \tag{3}$$

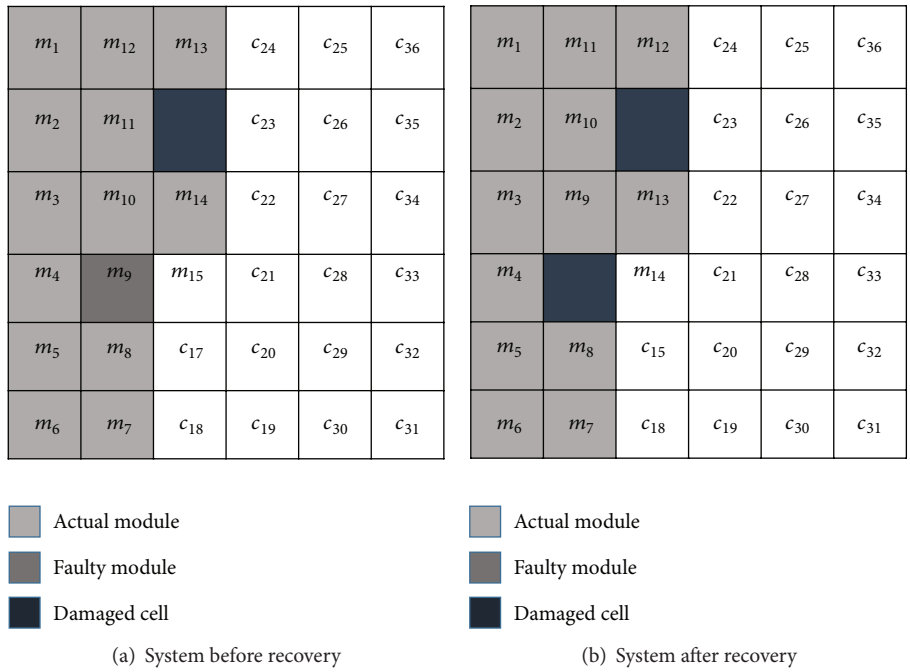
As SEEs arrive randomly, the fault occurrence process can be modeled by a Poisson process as shown in (4):

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \tag{4}$$

In (4), $P_k(t)$ is the fault occurrence probability, λ is the Poisson rate, t is the observation time and k is the fault occurrence times. For a normal SEE counting, $\lambda = 1$. Considering our fault-tolerance size ratio and fault-free nonoccupied CLBs in the FPGA, λ should be

$$\lambda = \frac{A_f}{S} = \frac{n(a + f)}{S}. \tag{5}$$

In the above equation, t is the observation cycle T. Because k is unknown, it is hard to obtain the fault occurrence



(a) System before recovery (b) System after recovery

Figure 5: Illustration of our fault recovery mechanism.

probability. However, it is easy to get the nonfault occurrence probability while $k = 0$; that is,

$$P_0 = e^{-(n(a+f)/S)T}. \quad (6)$$

Then the fault occurrence probability can be calculated as follows:

$$P_f = 1 - P_0 = 1 - e^{-(n(a+f)/S)T}. \quad (7)$$

In one observation cycle T, suppose there are M permanent faults. Then, the expected fault occurrence times are

$$n_{f_{EX}} = \sum_m P_f \times 1 = M \times \left(1 - e^{-(n(a+f)/S)T}\right). \quad (8)$$

The expected time interval between each pair of faults or expected system running time between each pair of faults is

$$t_{EX} = \frac{T}{n_{f_{EX}} - 1} = \frac{T}{M \times \left(1 - e^{-(n(a+f)/S)T}\right) - 1}. \quad (9)$$

With our fault recovery mechanism, the system recoverable times are also the number of faults which occurred before the system fails. Therefore, the expected system lifetime can be obtained by summing up all the expected intervals between each pair of faults as shown below:

$$T_{EX} = \sum_N t_{EX} = \frac{\left(\left(S - n\left(a + f\right)/\left(a + f\right)\right) + 1\right) \times T}{M \times \left(1 - e^{-(n(a+f)/S)T}\right) + 1}. \quad (10)$$

In order to find the optimal partition, we calculate all expected system lifetimes with all possible module size numbers from (10). Because the module size a should be an integer, we find all possible integers for module size partition and obtain the maximal T_{MAX} and its corresponding optimal partition. We propose Algorithm 1 to select the integer module size partition and find the maximal T_{MAX} and its corresponding optimal partition.

Input: All possible module partition numbers n $(1 \le n \le A)$ for an application whose size is A; FPGA size S.
Output: The maximal lifetime T_{MAX} and the optimal configuration.
(1) **if** $A + f \le S$
(2) $j \leftarrow 0$
(3) **for** $i = 1$ to A **do**
(4) **if** $S - (A + f \times i) \ge 0$ **and** $\lfloor A/i \rfloor = A/i$ **then**
(5) $a = A/i$
(6) Calculate $T_{EX}(i)$ with a using (10).
(7) $T_{MAX} = \mathbf{max}\left(T_{EX}(i)\right)$.
(8) $j = i$
(9) **end if**
(10) **end for**
(11) **end if**

Algorithm 1: Obtain the maximum expected system lifetime T_{MAX} and its corresponding partition.

Algorithm 1 presents a method to find the maximal lifetime T_{MAX} and its optimal configuration by using (10). $S - (A + f \times n(i)) \ge 0$ is used to guarantee that the total size with the fault detection modules is less than the size of the FPGA. $\lfloor A/n(i) \rfloor = A/n(i)$ is used to guarantee that it is an integer partition.

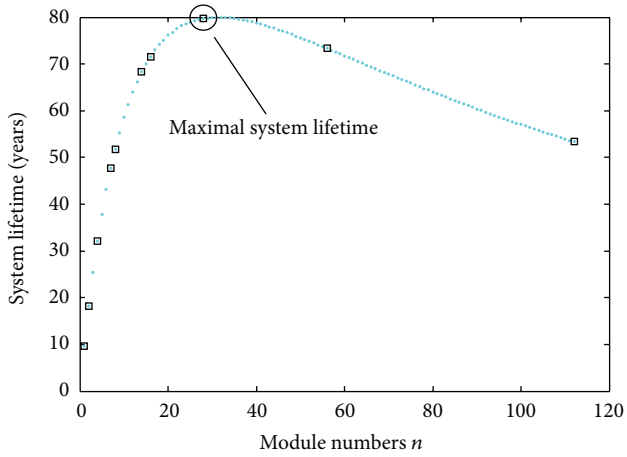


Figure 6: Illustration of searching optimal maximal system lifetime.

Table 1: System lifetime comparison between optimal partition and other partitions for SpaceWire.

Partitions ($a \times n$)	Expected system lifetime (years)
Optimal partition	
4×28	79.83
Other partitions	
1×112	53.36
2×56	73.31
7×16	71.56
8×14	68.26
14×8	51.72
16×7	47.70
28×4	32.04
56×2	18.19
112×1	9.57

We use SpaceWire, a commonly used communication protocol in spacecrafts, to illustrate the work flow of our technique. SpaceWire has 112 CLBs. We set the observation cycle T as one year, the fault ratio M as 122 times per year, and the FPGA size as 3424 CLBs. First, we input every module number n ($n = 1 \sim 122$) into (10) and let $a = A/n$. Hence, we get 122 calculated expected lifetimes from (10) with 122 different module numbers, as shown in Figure 6. Because a module size a should be an integer, by Algorithm 1, we select the integer partitions from those 122 outcomes. The curve in Figure 6 shows all integers obtained and corresponding to expected lifetimes. In Table 1, we compare the system lifetime from the optimal partition and some other partitions.

5. Experiments

In this section, we present and analyze the experimental results. We first introduce the experimental environment. Then we present the lifetime metric and measurement method. Finally, the experimental results are given and discussed.

We evaluate our system lifetime model on the XC2VP30 Xilinx XUP FPGA platform. We conduct experiments with six real applications: SpaceWire communication node (from European Space Agency), Wavelet image compressing, AC97 audio encoding and decoding (a modified demo from Xilinx), MEPG-4 video decoding (a modified demo from Xilinx), 8086 IP core, and Ethernet web service (a demo from Xilinx). We obtain the system lifetime of each of these six applications with the optimal partition. Additionally, we test three other different partitions for each application.

5.1. Experimental Environment. In our experiment, Xilinx XUP Virtex-II Pro development kit had been used as our validation platform, which has an FPGA chip XC2VP30-FFG896 with 3424 CLBs, and we use Matlab as a generator for fault simulation. Our layout design and modification tool is PlanAhead 9.2.7. For permanent fault simulation, we generate a permanent fault sequence with the Poisson process. The data for permanent faults simulation comes from NAOC (National Astronomical Observatories, Chinese Academy of Sciences), which are 122 permanent faults per year averagely at 14 MeV neutron and 85°C.

5.2. Performance Metrics and Measurement Method. Permanent faults by SEE are counted by years [29, 30]. To speed up our experiments, we simulate one year's permanent faults in one minute. The faults occurring in FPGA follow the discrete uniform random distribution. Therefore, we generate a discrete uniform random sequence for fault location simulation and the range of the sequence is from the minimal SN to the maximal SN of cells. To simulate permanent faults in FPGA, we place an empty module in a damaged cell, so the logic resource of this cell cannot be used anymore.

In our experiments, we divide an FPGA to cells with the same size and sequentially mark cells (see Section 3) and then configure the initial layout with the partition (including fault detection modules) of an application. SEE-induced permanent faults are generated as described above. Once a simulated fault occurs, the fault recovery mechanism works for fault recovery. After every reconfiguration, we check if there is any routing problem in the new layout and if the system is running well. Furthermore, we record the running time of each application.

5.3. Results and Discussion. Six different applications, SpaceWire (112 CLBs), Wavelet (198 CLBs), AC97 (350 CLBs), MEPG-4 (544 CLBs), 8086 (855 CLBs), and Ethernet (1680 CLBs), are used in our experiments.

5.3.1. Comparison between Testing Results and Expected Results with Optimal Partition. We tested each system 100 times with respective optimal partition calculated by (10) and Algorithm 1, and the test results are shown in Figure 7 and Table 2.

As it is shown in Figure 7, each line is the maximum expected system lifetime from (10); the marked points around each expected result are the test results from the 100 times experiments. In Table 3, we present a comparison between

Table 2: Comparison of the system lifetime between optimal partition and other three partitions for each application.

Application system	Partition ($a \times n$) (CLBs)		Expected lifetime (years)	Average test results (years)	Difference (years)
SpaceWire	Optimal partition	4×28	79.83	79.93	0.2831
	Tile-based partition	1×112	53.36	53.28	0.2121
	Optional partition	28×4	32.04	32.05	0.3231
	Nonpartition	112×1	9.57	9.52	0.2815
Wavelet	Optimal partition	3×66	40.98	40.94	0.2774
	Tile-based partition	1×198	27.24	27.32	0.3093
	Optional partition	66×3	7.63	7.68	0.2526
	Nonpartition	198×1	2.85	2.94	0.2667
AC97	Optimal partition	5×70	20.62	20.63	0.2395
	Tile-based partition	1×350	12.69	12.65	0.2727
	Optional partition	70×5	3.71	3.69	0.2489
	Nonpartition	350×1	0.82	0.81	0.1738
MEPG-4	Optimal partition	4×136	12.37	12.38	0.2038
	Tile-based partition	1×544	5.55	5.56	0.2329
	Optional partition	136×4	1.22	1.18	0.2466
	Nonpartition	544×1	0.35	0.47	0.1247
8086	Optimal partition	5×171	6.58	6.56	0.2575
	Tile-based partition	1×855	0.03	0.04	0.0063
	Optional partition	171×5	0.57	0.56	0.0326
	Nonpartition	855×1	0.12	0.13	0.0072
Ethernet	Optimal partition	8×210	1.73	1.71	0.1182
	Tile-based partition	N/A	N/A	N/A	N/A
	Optional partition	420×4	0.11	0.12	0.0820
	Nonpartition	1680×1	0.04	0.04	0.0012

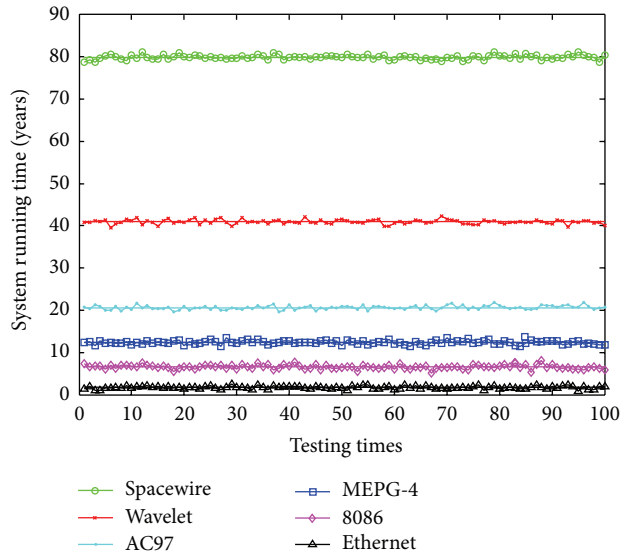


Figure 7: Test results of each application compared with expected results.

the expected lifetime and the average test result of each application.

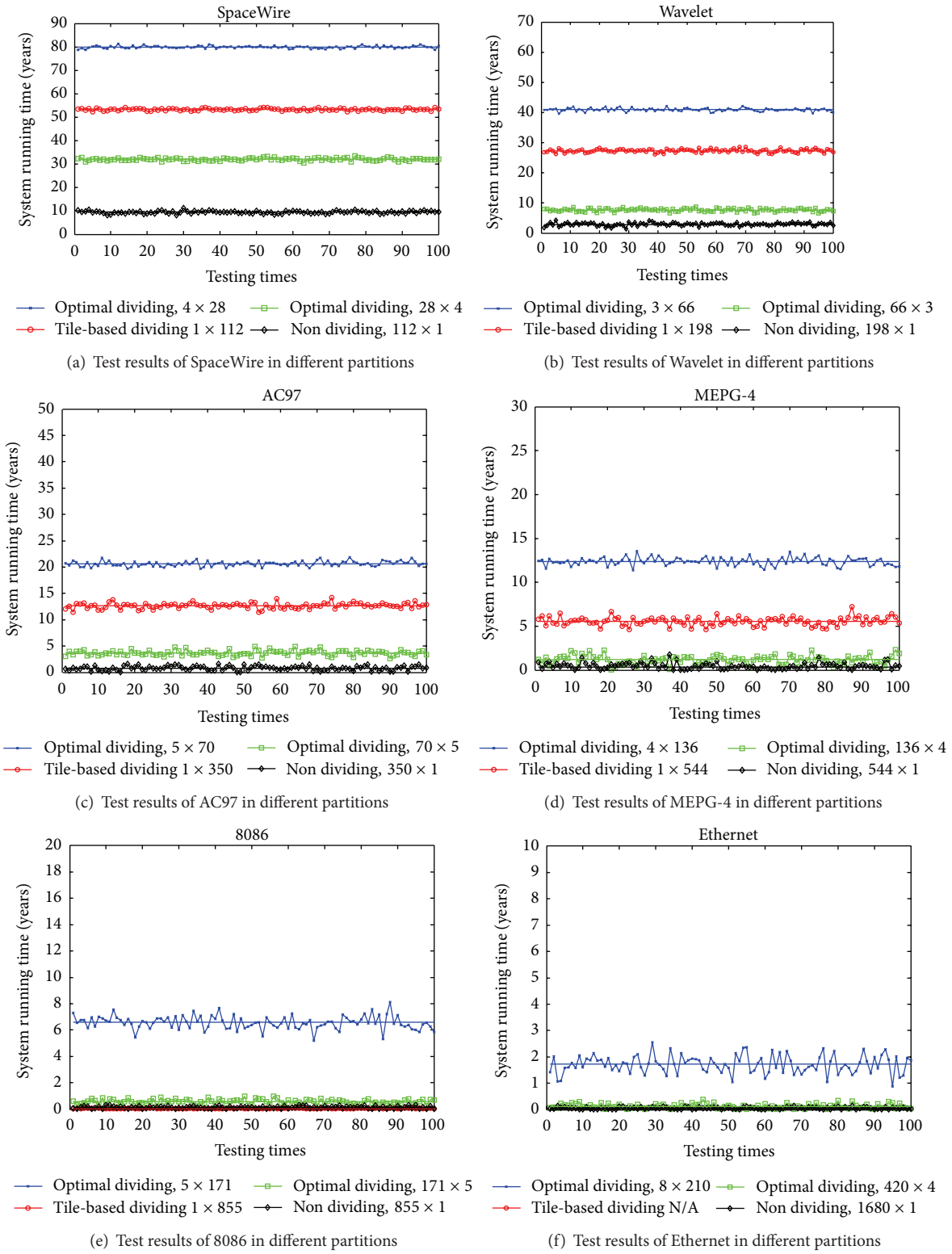
In Figure 7, each application's test results are distributed closely around their expected lifetime. Table 3 only shows little difference between the average test results and expected lifetime.

Table 3: Comparison of the system lifetime between the expected and test results in optimal partition.

Application	Optimal partition ($a \times n$) (CLBs)	Expected results (years)	Average test results (years)
SpaceWire	4×28	79.83	79.93
Wavelet	3×66	40.98	40.94
AC97	5×70	20.62	20.63
MEPG-4	4×136	12.37	12.38
8086	5×171	6.58	6.56
Ethernet	8×210	1.73	1.71

By using our technique, although some small-size applications (i.e., SpaceWire and Wavelet) have very long lifetime that may exceed the spacecraft's lifetime for current space mission, in the real design, we can use a small-size FPGA for these applications to reduce unnecessary lifetime and save hardware cost.

5.3.2. System Lifetime Comparison between Optimal Partition and Other Partitions. Furthermore, we test other three partitions for each system: tile-based partition (use 1 CLB as function module for fault recovery, $a = 1$) [16, 18, 19, 31], optional partition (partition application into 3 to 5 modules,



(a) Test results of SpaceWire in different partitions

(b) Test results of Wavelet in different partitions

(c) Test results of AC97 in different partitions

(d) Test results of MEPG-4 in different partitions

(e) Test results of 8086 in different partitions

(f) Test results of Ethernet in different partitions

Figure 8: Test results of optimal partition compared to other three partitions for each application.

Table 4: System lifetime improvements by using optimal partition.

Comparative partition	Lifetime improvements (%)
Tile-based partition	3362.12
Optional partition	733.87
Non-partition	2700.58

$n = 3 \sim 5$), and nonpartition (i.e., using the whole application as a reconfiguring module, $a = A$). We compare the experimental results of these partitions with those of the optimal partition. The results are shown in Figure 8.

In every subfigure of Figure 8, every line is the expected lifetime of each different partition from (10); the marked points around each expected result are the test results from the 100 times experiment with different partitions. For each application in Table 4, we present a comparison between the expected lifetime and the average lifetime for each partition.

In Figures 8(d), 8(e), and 8(f), because the system lifetimes from the nonoptimal partitions are too close to each other, they overlap each other. For the application Ethernet, with tile-based partition, its space requirement exceeds the whole size of FPGA ($(1 + 3) \times 1680 = 6720 > 3424$), so we did not test this case. Table 3 shows a good consistency between the expected lifetime and the average results even for the nonoptimal partition, which proves the correctness of our system lifetime model.

Compared with other three partitions, our optimal partition always has the better performance in terms of the system lifetime for each application. By using our optimal partition, for each application, the average improvements (with six applications) of the system lifetime over the three partitions are shown in Table 4.

The experimental results show that the optimal partition from our model can have better system lifetime than other partitions in every application. We did not find any routing failure problems during the experiments, and the system performance degradation after each reconfiguration is acceptable.

6. Conclusion

In this paper, we proposed a system lifetime estimation model for SRAM-based FPGAs, which can find the optimal partition to maximize the system lifetime for space mission. We developed a permanent fault recovery mechanism to recover permanent faults by partial reconfiguration. We built a lifetime model to obtain the maximal system lifetime with its corresponding optimal partition. The experiment results show that our technique can effectively improve the system lifetime.

References

[1] J. T. Wallmark and S. M. Marcus, "Minimum size and maximum packing density of nonredundant semiconductor devices," in *Proceedings of Institute of Radio Engineers*, pp. 286–298, Institution of Radio Engineers, 1962.

[2] D. Binder, E. C. Smith, and A. B. Holman, "Satellite anomalies from galactic cosmic rays," *IEEE Transactions on Nuclear Science*, vol. 22, no. 2, pp. 2675–2680, 1975.

[3] D. Weigand and M. Harlacher, "A radiation-tolerant low-power transceiver design for reconfigurable applications," in *Proceedings of the Earth Science Technology Conference*, p. A1P2, 2002.

[4] K. Morris, "SFPGAs in space," Tech. Rep., FPGA and Structured ASIC Journal, 2004.

[5] D. Ratter, "FPGA on mars," Tech. Rep., xCell Journal Xilinx, 2004.

[6] A. Dawood, S. Visser, and J. Williams, "Reconfigurable FPGAs for real time image processing in space," in *Proceedings of the 14th International Conference on Digital Signal Processing (DSP '02)*, July 2002.

[7] P. Bergsman, "Xilinx fpga blasted into orbit," Tech. Rep., xCell Journal Xilinx, 2003.

[8] E. Fuller, M. Caffrey, A. Salazar, C. Carmichael, and J. Fabula, "Radiation characterization and seu mitigation of the virtex fpga for space-based reconfigurable computing," in *Proceedings of the IEEE Nuclear and Space Radiation Effects Conference*, 2000.

[9] D. Lim and M. Peattie, *A Space-Efficient Flash Translation Layer for CompactFlash Systems*, vol. 290, Xilinx Application Notes, 2004.

[10] M. Berg, C. Poivey, D. Petrick et al., "Effectiveness of internal versus external SEU scrubbing mitigation strategies in a Xilinx FPGA: sesign, test, and analysis," *IEEE Transactions on Nuclear Science*, vol. 55, no. 4, pp. 2259–2266, 2008.

[11] C. Bolchini, D. Quarta, and M. D. Santambrogio, "SEU mitigation for sram-based fpgas through dynamic partial reconfiguration," in *Proceedings of the 17th Great Lakes Symposium on VLSI (GLSVLSI '07)*, pp. 55–60, Maggiore, Italy, March 2007.

[12] K. Nakahara, S. Kouyama, T. Izumi, H. Ochi, and Y. Nakamura, "Fault tolerant reconfigurable device based on autonomous-repair cells," in *Proceedings of the International Conference on Field Programmable Logic and Applications (FPL '06)*, pp. 1–6, August 2006.

[13] M. Abramovici, C. Stroud, B. Skaggs, and J. Emmert, "Improving online bist-based diagnosis for roving stars," in *Proceedings of the 6th IEEE International Online Testing Workshop (IOLTW '00)*, p. 31, Mallorca, Spain, 2000.

[14] J. Emmert, C. Stroud, B. Skaggs, M. Abramovici et al., "Dynamic fault tolerance in FPGAs via partial reconfiguration," in *Proceedings of the IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM '00)*, p. 165, 2000.

[15] J. M. Emmert, C. E. Stroud, and M. Abramovici, "Online fault tolerance for FPGA logic blocks," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 15, no. 2, pp. 216–226, 2007.

[16] S. Pontarelli, M. Ottavi, V. Vankamamidi, G. C. Cardarilli, F. Lombardi, and A. Salsano, "Analysis and evaluations of reliability of reconfigurable FPGAs," *Journal of Electronic Testing*, vol. 24, no. 1–3, pp. 105–116, 2008.

[17] A. Kanamaru, H. Kawai, Y. Yamaguchi, and M. Yasunaga, "Tile-based fault tolerant approach using partial reconfiguration," in *Proceedings of the 5th International Workshop on Reconfigurable Computing: Architectures, Tools and Applications (ARC '09)*, pp. 293–299, 2009.

[18] J. Lach, W. H. Mangione-Smith, and M. Potkonjak, "A space-efficient flash translation layer for CompactFlash systems," *IEEE Transactions on Consumer Electronics*, vol. 48, no. 2, pp. 366–375, 2002.

[19] A. Antola, M. Sami, and V. Piuri, "Online diagnosis and reconfiguration of fpga systems," in *Proceedings of the 1st IEEE International Workshop on Electronic Design, Test and Applications (DELTA '02)*, p. 291, 2002.

[20] W.-J. Huang and E. J. McCluskey, "Column-based precompiled configuration techniques for fpga," in *Proceedings of the 9th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM '01)*, pp. 137–146, 2001.

[21] K. Paulsson, M. Hübner, M. Jung, and J. Becker, "Methods for run-time failure recognition and recovery in dynamic and partial reconfigurable systems based on xilinx virtex-II pro FPGAs," in *Proceedings of the IEEE Computer Society Annual Symposium on Emerging VLSI Technologies and Architectures (ISVLSI '06)*, pp. 159–164, March 2006.

[22] F. Lima, L. Cairo, and R. Reis, "Designing fault tolerant systems into SRAM-based FPGAs," in *Proceedings of the 40th Design Automation Conference (DAC '03)*, pp. 650–655, June 2003.

[23] L. Sterpone and M. Violante, "A new reliability-oriented place and route algorithm for SRAM-based FPGAs," *IEEE Transactions on Computers*, vol. 55, no. 6, pp. 732–744, 2006.

[24] F. Lima, C. Carmichael, J. Fabula, R. Padovani, and R. Reis, "A fault injection analysis of virtex FPGA TMR design methodology," in *Proceedings of the 6th European Conference on Radiation and Its Effects on Components and Systems*, pp. 275–282, September 2001.

[25] C. Carmichael, *Triple Module Redundancy Design Techniques for Virtex fpgas*, vol. 197, Xilinx Application Notes, 2006.

[26] D. R. Czajkowski, P. K. Samudrala, and M. P. Pagey, "SEU mitigation for reconfigurable FPGAs," in *Proceedings of the IEEE Aerospace Conference*, p. 7, 2006.

[27] F. L. Kastensmidt, L. Sterpone, L. Carro, and M. S. Reorda, "On the optimal design of triple modular redundancy logic for SRAM-based FPGAs," in *Proceedings of the Conference on Design, Automation and Test in Europe (DATE '05)*, pp. 1290–1295, March 2005.

[28] M. Wirthlin, E. Johnson, N. Rollins, M. Caffrey, and P. Graham, "The reliability of FPGA circuit designs in the presence of radiation induced configuration upsets," in *Proceedings of the 11th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM '03)*, p. 133, 2003.

[29] P. Layton, E. Patnaude, G. Williamson, L. Longden, and C. Sloan, "Compendia of radiation test results of integrated circuits," in *Proceedings of the IEEE Radiation Effects Data Workshop*, pp. 156–162, July 2005.

[30] W. E. Willing and N. P. Goldstein, "Combining single-event latchup and reliability requirements for space vehicles," in *Proceedings of the Annual Reliability and Maintainability Symposium*, pp. 445–449, January 1995.

[31] S. Pontarelli, G. C. Cardarilli, A. Malvoni, M. Ottavi, M. Re, and A. Salsano, "System-on-chip oriented fault-tolerant sequential systems implementation methodology," in *Proceedings of the IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT '01)*, pp. 455–460, October 2001.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 682085, 8 pages
http://dx.doi.org/10.1155/2013/682085

Research Article

FPGA Implementation of Real-Time Ethernet for Motion Control

Chen Youdong,[1] Xing Chunxiang,[1] Tao Yong,[1] and Sun Kai[2]

[1] *School of Mechanical and Automation, Beihang University, Beijing 100191, China*
[2] *College of Mechanical and Electrical Engineering, North China University of Technology, Beijing 10041, China*

Correspondence should be addressed to Chen Youdong; chenyd@buaa.edu.cn

Received 22 April 2013; Accepted 12 June 2013

Academic Editor: Shao Zili

This paper provides an applicable implementation of real-time Ethernet named CASNET, which modifies the Ethernet medium access control (MAC) to achieve the real-time requirement for motion control. CASNET is the communication protocol used for motion control system. Verilog hardware description language (VHDL) has been used in the MAC logic design. The designed MAC serves as one of the intellectual properties (IPs) and is applicable to various industrial controllers. The interface of the physical layer is RJ45. The other layers have been implemented by using C programs. The real-time Ethernet has been implemented by using field programmable gate array (FPGA) technology and the proposed solution has been tested through the cycle time, synchronization accuracy, and Wireshark testing.

1. Introduction

Motion control systems have been widely used in various applications of factory automation systems for a long time. The basic motion control system component includes controller, servo drivers, and motors. Now, more and more controllers connected servo drivers use the communication network. Actually, many types of networks having been used ranging from proprietary connections to open connections in motion control systems. The communication systems in the motion control system should guarantee the domain-specific requirements: real-time (maximum transfer delay, jitter in the transmissions, and available bandwidth).

Ethernet is so fast, easy to install, and cheap that it wins widespread popularity. However, Ethernet is not a real-time network and thus may cause time delay. Ethernet can be effectively applied in real-time communication with modification. There are different ways of modifying the Ethernet technology [1]. All the solutions proposed can be classified into three different approaches in principle: on top of TCP/IP, on top of Ethernet, and modified Ethernet [2].

Within the automation domain, the real-time requirements focus on the response time behavior of data packets. There are three real-time classes to guarantee response time: soft real-time, hard real-time, and isochronous real-time [3]. Only isochronous real-time Ethernet is used for motion control. The isochronous real-time Ethernet (RTE) devices provide more predictable and reliable real-time data transfer and means to support the precise synchronization of automation equipment according to IEEE 1588 [4, 5]. Several isochronous RTE devices are available from different vendors, such as Powerlink [6], Ethercat [7], Sercos III [8], and Profinet IO [9]. Most of them comply with both the IEC 61158 [10] and IEC 61784-2 [11] International Standards.

Field Programmable Gate Arrays (FPGAs) are semiconductor devices that are based around a matrix configurable logic blocks (CLBs) connected via programmable interconnects. All the internal logic elements and all the control procedures of the FPGA are executed continuously and simultaneously. Therefore, the execution time of FPGA is faster than either the digital signal processor (DSP) or the personal computer (PC). Recently, some researches adopted the FPGA chip to implement low-cost, high-performance, real-time industrial applications [12, 13].

The protocols of available RTE network are too complex to grasp and thus add to difficulty in its development. They use specific chips and communication cards, another negative element on road of its development. This paper proposes an

isochronous RTE network named CASNET, which modifies the Ethernet MAC achieved by the FPGA to meet the real-time requirements for motion control. CASNET uses a ring topology within the segment. The MAC adopts the Master/Slave principle, where Master node (typically the control system) sends the Ethernet frames to the Slave nodes which extract data from and insert data into these frames. The physical layer is based on the Ethernet hardware.

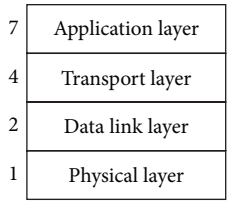


Figure 1: CASNET protocol layer structure.

2. CASNET Protocol

CASNET adopts simple three layers for a simple configuration proposed by IEC, as opposed to selecting the hierarchical model suggested by the OSI-7 layer [14, 15]. However, it appends on the fourth layer, referred to as the transport layer, to package and send/receive the message in CASNET device, as shown in Figure 1. The data link layer of CASNET protocol works in hardware. VHDL is used as an implementation and the functional verification is carried out by using FPGA.

Service provided in each layer is explained as follows.

The communication media is connected directly to the physical layer, where both status of the media and synchronization are controlled. The interface of the physical layer is Ethernet (RJ45).

The data link layer is designed for making the physical link reliable. This is achieved by providing CRC check. The link layer sends and receives telegrams named as CASNET frames as shown in Figure 2. Upon arrival of the frame, the data link layer receives incoming telegram, extracts the relevant user data, sends the user data to Transport layer, inserts the slave station relevant data to the telegram, and transfers the telegram to the next CASNET slave station. Figure 2 also shows the structure of the CASNET telegrams contained in the Ethernet data field. A telegram begins with a 1-byte address that represents the area in which slaves write or read user data. The address is followed by a 1-byte control word that specifies the type of control or operation for which it is intended (e.g., position, velocity, write, and read). A 1-byte status represents the telegram from/to master station. A 1-byte time is used to record the time stamp to synchronize clock. A 6-byte periodic data and a 2-byte aperiodic data are real-time data exchanging between master and slave.

The transport layer is not only for packaging the application data into frames and transmitting the data to the link layer, but also for reassembling telegrams coming from the link layer. In CASNET, the transport layer is incorporated into the application layer. The application layer cooperates with the transport layer and the data link layer to enable reliable communications. It provides standardized functions and data formatting with which the user can interact.

CASNET is a master/slave network that composes a master station and up to 125 slave stations. The system topology structure is shown in Figure 3. The CASNET network is a ring topology achieved by point-to-point connections between consecutive nodes. The network controllers (either master or slaves) are full-duplex devices that are capable of receiving/transmitting data concurrently. CASNET adopts standard Ethernet frames that encapsulate the telegrams specifically defined by the protocol, as shown in Figure 2. The master station is a standard Ethernet interface running the CASNET protocol. The slave consisted of the standard Ethernet interface and real-time MAC achieved by FPGA.

The master station may be CNC controller or industrial robot controller and it packages data, analyzes data, initializes systems, data communication control, synchronizing clock, and sends data package.

At the runtime, a single frame periodically issued by the master station circulates among all the slave stations. On the arrival of the frame, each slave station receives incoming telegrams, extracts the relevant user data, sends the user data to the controller relevant to the slave, inserts the slave station-related data to the telegram, and transfers the telegram to the next CASNET slave station through Ethernet Rx wire pairs. The last CASNET slave station sends the fully processed telegram back through Ethernet Tx wire pairs.

3. CASNET Slave Controller

Both the CASNET master and the slave can in principle be operated by using Ethernet physical layer, such as copper cables (100 base TX) and fiber optic (100 base Fx). The CASNET master is normally operated with standard network interfaces (in this context referring to any PC compatible MAC and PHY). However, the CASNET slave is operated by special hardware (e.g., FPGA) to facilitate very short packet forwarding delays in the slave devices.

The slave controller is operated for the data link layer by VHDL in a CASNET protocol. The other layers in the slave controller are operated by using software programs. A functional test is performed by FPGA. Firmware and test board are developed to examine operations of CASNET protocol designed with FPGA, as shown in Figure 4.

Figure 5 depicts an overall block diagram of the designed CASNET slave controller. The CASNET slave controller proposed in this paper mainly consists of three blocks such as the physical layer (PHY), the real-time media access control (MAC), and microcontroller. The data interface between the real-time MAC and the micro-processor, the interface between the real-time MAC and the PHY, are implemented in the real-time MAC. As the microprocessor and PHYs adopt the standard chip, these blocks are not dealt with in this paper.

3.1. CASNET Slave Controller Registers. The CASNET slave controller has an address space of 4 kByte. The memory space in the real-time MAC is used for registers and user memory. The registers are configured by the CASNET telegram at

Table 1: CASNET slave register and user memory address.

Register	Address (hex)	CPU R/W	Casnet R/W	Description
Trans_buf (ram)	0x0~0x1ff	RW	RO	Slave data buffer that is to be transmitted
Rev_buf (ram)	0x200~0x3ff	RO	WO	Slave receiving data buffer
CMD	0x400~0x401	RO	RW	The salve command of the current frame
ADDR	0x402~0x403	RO	RW	The address that slave receives data from the current frame.
OFFS	0x404~0x405	RO	RW	The slave offset address of the current frame
LEN	0x406~0x407	RO	RW	The current frame data length
INT	0x40a	RW	—	Interrupt register
DTS0	0x410~0x417	RO	RO	Local time when receiving frame on Port 0
DTS1	0x418~0x41f	RO	RO	Local time when receiving frame on Port 1
LTIME	0x420~0x427	RO	RO	Local time
MTIMECOPY	0x428~0x42f	RO	RW	Local time of system time.
DELAY	0x430~0x437	RO	RW	Propagation delay between slave and master
OFFTIME	0x438~0x43f	RO	RW	Difference between local time and system time
DIFFTIME	0x440~0x448	RO	RO	Mean time: the mean of the latest eight times clock drift
REFTIME	0x450~0x457	RO	RO	Reference time
ID	0x7fe~0x7ff	RO	RW	Slave ID

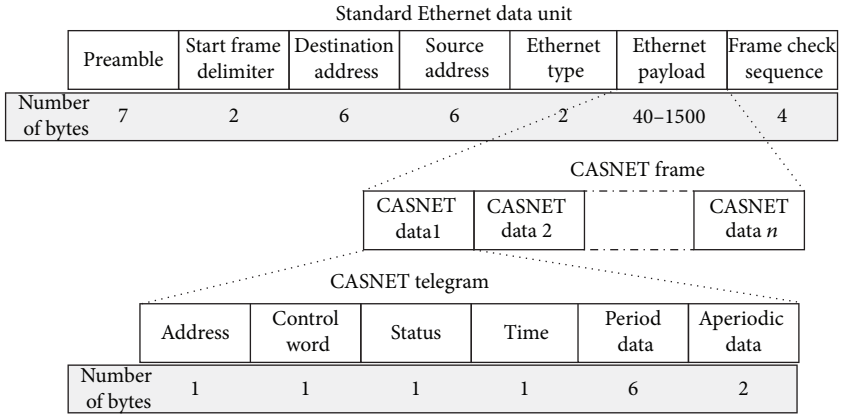


Figure 2: Structure of CASNET frame.

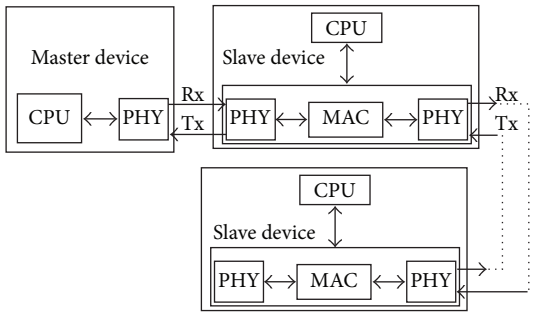


Figure 3: System topology structure.

Figure 4: Board with FPGA implementing CASNET slave.

initialization. The address range is directly addressable by the CASNET master and an attached μC_If block. The μC_If block realizes the connection between the real-time MAC and the microcontroller. Table 1 is an overview of the registers and user memory.

3.2. Internal Configuration of Real-Time Media Access Control (MAC). The architecture of the real-time MAC is shown in Figure 5. The real-time MAC is composed of six units, namely, RMII_mux, CASNET receiving block (Casnet_rev), CASNET

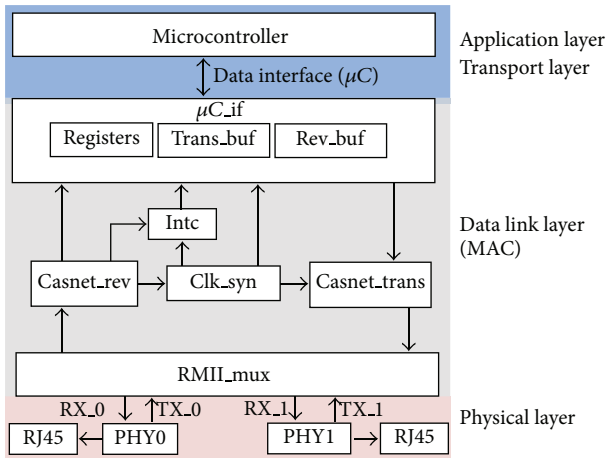


Figure 5: Overall block diagram of CASNET slave controller. RMII_mux—reduced media independent interface; Casnet_rev—CASNET receive; Clk_syn—clock synchronization; Intc—interrupt controller; μC_if—Microcontroller interface; Casnet_trans—CASNET transmit; Trans_buf—Slave data buffer that is to be transmitted; Rev_buf—Slave receiving data buffer.

transmitting block (Casnet_trans), clock synchronization block (Clk_syn), interrupt control (Intc), and microcontroller interface (μC_if). The MAC works as follows: when a frame is received through the RMII_mux (Tx and Rx lines), the Casnet_rev receives and extracts the relevant slave user data, then writes the user data into the Rev_buf ram block of the μC_if. The μC_if is a communication interface between micro-processor and the real-time MAC. The microcontroller accesses the user data in the Rev_buf ram of the μC_if and writes the return data (the slave status, warning, local timer, etc.) into the Trans_buf ram block of the μC_if. The Casnet_trans reads the return data from the Trans_buf ram, inserts the return data into the frame, and forwards this frame to the next slave. If the user data received by the Casnet_rev is synchronization data, this synchronization data is written to the Clk_syn unit and microcontroller to align the distributed clock and adjust the local timer. When the user data or synchronization data is completely received, the Intc unit generates interrupt to remind the microcontroller respectively. The following sections describe the functions and operations related to each block in more detail.

3.2.1. RMII_Mux. The RMII_mux is used to connect the real-time MAC with the Ethernet PHYs. It supports 100 MB/s data rates. The RMII_mux unit detects the connection between CASNET slave and PHYs. If the CASNET slave connects the PHYs, the RMII_mux unit connects the signals Rx0 and Tx0 of the real-time MAC to the Rx0 and Tx0 of the PHY0, respectively, and the signals Rx1 and Tx1 of the real-time MAC to the Rx1 and Tx1 of the PHY1, respectively, otherwise it can connect Rx to Tx to form a loop in the real-time MAC.

3.2.2. Receiving and Transmitting Block. The internal structure of receiving unit Casnet_rev is represented in Figure 6(a). The finite state machine (FSM) controls the operation when user data is being extracted from the frame. Once the receiving user data is completed, a RCF signal writes to Intc unit. As soon as the Intc unit receives the RCF signal, it immediately generates an interrupt to remind the microcontroller. If the user data is synchronization data, it is forwarded to Clk_syn unit via Clk_syn_if to synchronize the distributed clock. CLK_syn_if is a 64 bit read/write interface. The user data received is stored in FIFO. The FIFO size is set by parameters, and its maximum size is up to 32×16 bit. The frame of slave is up to 2 kbit. After the user data stored in the receiving FIFO pass the Cyclic Redundancy Check of 32 bits (CRC-32), they are sent to the Rev_buf ram of μC_if unit.

The internal structure of transmitting unit Casnet_trans is given in Figure 6(b). If there is data transmitting, the user data that pass CRC-32 check will be sent from the Trans_buf ram of μC_if unit to the transmitting FIFO. The FSM controls the transmitting process.

The CRC-32 polynomial is shown in (1). The CRC calculation can be realized with a shift register and exclusive-or (XOR) logic gates, where the data is processed one bit per clock cycle. There are necessary $(k + r)$ cycles to obtain an r-bit CRC code for a message with k bits [15]. Although the serial implementation is simple and can run at a high clock rate, it is low data throughput for its serial input. To increase the throughput of CRC calculation, the parallel implementation is used to calculate and transmit it, which is adopted by [16]

$$p(x) = x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^8 + x^7 + x^5 + x^4 + x^2 + x + 1. \quad (1)$$

Considering the CASNET protocol, the FSM of receiving and transmitting data is adopted as shown in Figure 7. Once there is the RMII input, the state machine will switch to the Header state from the IDLE state. The Header state checks the frame header to determine whether it is a CASNET message. The state will switch to the ADDRESS state after the CASNET frame passes check, or it will switch to the ERROR state. The CASNET command or control types are received/transmitted in the ControlWord state. The command types are read or write operation, the read operation is performed before the write operation. The control types include position mode, velocity mode, acceleration mode, and IO mode. The status is read to determine the telegram from/to master station in the STATUS state. The operation of DATA state abstracts or inserts user data from/to CASNET frame at the address space. If the data is received, the user data will be stored in the receiving FIFO. The state will switch to the NOTIFY state after the user data pass the CRC check. If the data are being received, the NOTIFY state will send out a signal to notify the Rev_buf ram to store the received data. If the data are transmitted, the NOTIFY state will give a signal to copy the data in the Trans_buf ram to the transmitting FIFO. The state switches to IDLE state after the data are completely stored in the Rev_buf ram or the transmitting FIFO. The conditions of FSM transition are depicted in Table 2.

3.2.3. Clock Synchronization Block. Clock synchronization block enables all CASNET devices (master and slaves) to

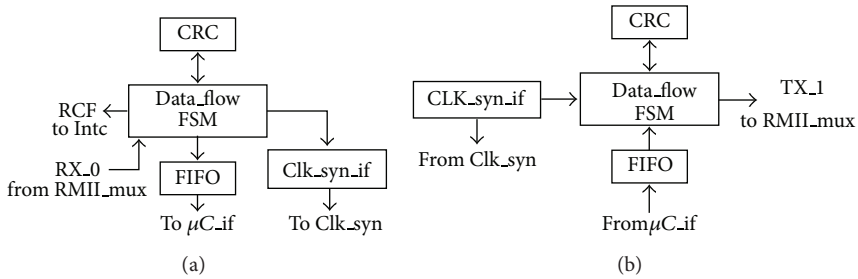


Figure 6: (a) Casnet_rev structure. (b) Casnet_trans structure. FSM—finite state machine, CRC—Cyclic Redundancy Check, RCF—receive complete flag. SWF—successful writes flag. FIFO—fist input first output. Clk_syn_if—Clk_syn interface.

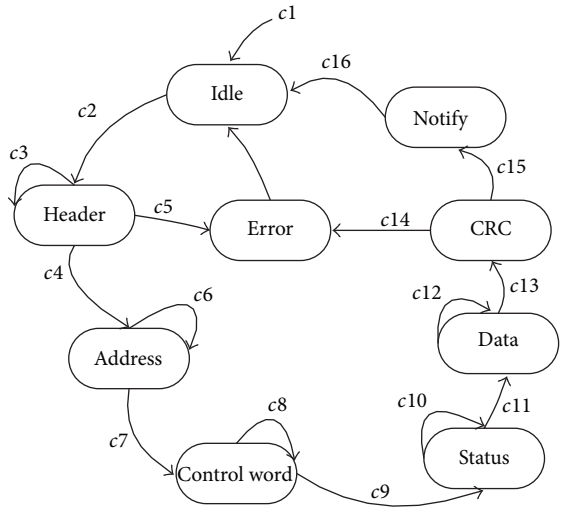


Figure 7: Finite state machine of receive/transmit.

Table 2: Conditions of FSM transition.

Conditions	Description
$c1$	There is no RMII input
$c2$	There is RMII input
$c3$	Check if it is CASNET header
$c4$	It is CASNET header
$c5$	It is not CASNET header
$c6$	Check whether address received/transmitted is complete
$c7$	The address received/transmitted is complete
$c8$	Check whether ControlWord received/transmitted is complete
$c9$	The ControlWord received/transmitted is complete
$c10$	Check whether Status received/transmitted is complete
$c11$	The Status received/transmitted is complete
$c12$	Check whether data received/transmitted is complete
$c13$	The data received/transmitted is complete
$c14$	It do not pass the CRC check
$c15$	It pass the CRC check
$c16$	The data store in FIFO

share the same time. It may be achieved either by hardware or software. The CASNET clock synchronization is hardware based on the master and slave clock, which is very closely related to the IEEE 1588 standard. A master clock of a specified device is used to synchronize the slave clock of the other devices. A synchronization telegram containing the current master clock time is sent at certain intervals and the devices with slave clocks read the time from the same telegram to synchronize the local time.

In this system, there are three types of compensation: propagation delay from the device of master clock to the devices of slave clock, clock drift caused by oscillator instabilities due to temperature changes, aging and other reasons, and the offset between local clock and master clock caused by the initial difference of the local times resulting from different times at which the system is powered up. The first two compensations should be compensated at certain intervals while the last can be compensated at start-up.

The clock synchronization process consists of three steps: propagation delay measurement, offset compensation, and clock drift compensation.

Propagation delay measurement is initiated by the master with time stamping between all slaves, which adopt a master-slave based protocol termed precision time protocol (PTP). The propagation delay t_{delay} is calculated by the device of the master clock and written to the devices of slave register DELAY.

The offset is calculated by the device of the master clock and is written to the devices of the slave register OFFTIME. Each device of slave calculates its local copy of the master time t_{time} using its local time t_{local} and local offset value t_{offset}:

$$t_{\text{time}} = t_{\text{local}} + t_{\text{offset}}. \tag{2}$$

After the propagation delay has been measured and the offset compensated, the clock drift Δt of every local clock is compensated by a clock rate algorithm, which is achieved by hardware. The $t_{\text{reference}}$ is the copy of the master clock time, which is stored at slave register REFTIME as follows:

$$\Delta t = t_{\text{time}} + t_{\text{delay}} - t_{\text{reference}}. \tag{3}$$

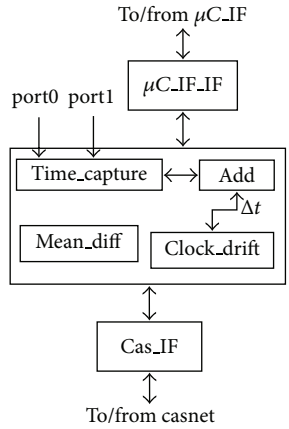


Figure 8: Structure of clock synchronization.

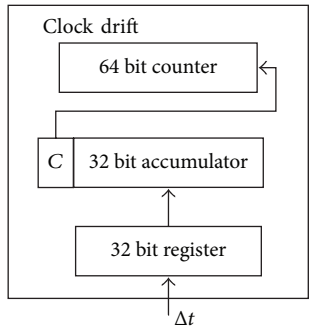


Figure 9: Structure of 40 M divider.

If Δt is positive, it means that the local clock is running faster than the master clock and thus needs to be slowed down; if Δt is negative, it means that the local time is running slower than the master clock and has to be sped up.

The clock synchronization unit includes timestamp capturing Time_Capture unit, clock drift compensation Clock_drift unit, adder Add, and clock drift averaging Mean_diff unit, as shown in Figure 8. The Add is used to calculate the clock drift Δt. The following sections describe the functions and operations related to each unit in clock synchronization block

(a) Timestamp Capturing. The timestamp is captured at the first shot of CASNET packet arriving at Rx or Tx (Rx_DV = 1) and writes the time to the local time registers DS0 and DS1.

(b) Clock Drift Compensation. Clock drift compensation can be achieved by employing an oscillator with small drift and clock rate algorithm. Unfortunately, accurate oscillators are very expensive and consume much more power and space. The implementation of clock rate algorithm maintains the clock rate stability and is much cheaper in term, of additional communication as well as computation cost.

A dedicated compensate clock IP implements the clock rate algorithm, as shown in Figure 9. Actually it is a clock counter that can fine tune the frequency of the calculator. It

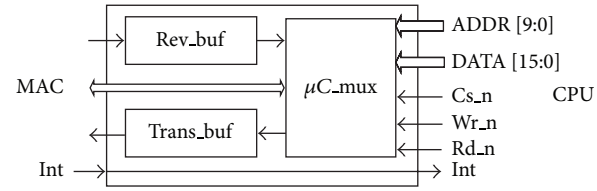


Figure 10: Structure of μC_if.

is composed of a 64 bit clock calculator, a 32 bit accumulator, and a 32 bit register. The accumulator is a divider. The register is used to store the compensating rate r_c. They are all driven by oscillator whose frequency is f_{osc}. Accumulator adds its value to compensating rate r_c at every oscillator period, and the accumulated value is stored to the accumulator. At the same time, a carry flag is generated to indicate whether the addition is overflow. If it is overflow, the clock calculator will get an increment value that is equal to the resolution of the compensate clock at the next oscillator period. The frequency f_{cal} of clock counter is determined by r_c and f_{osc} as follows:

$$f_{\text{cal}} = \frac{f_{\text{osc}} \times r_c}{2^{32}}. \tag{4}$$

When f_{osc} changes, r_c can be tuned to maintain f_{cal}. In this system, the frequency of oscillator f_{osc} is 50 MHZ, and the frequency f_{cal} of clock counter is 40 MHZ. the r_c can be obtained as follows:

$$\Delta r_c = \left(\Delta t_1 \ll 1 - \Delta t_0\right) \ll k, \tag{5}$$

where Δt_1 is the clock drift this time, Δt_0 is the clock drift last time, and default is $32'$b0, k is the damping coefficient, and k is 5. Δr_c may be negative.

The value of accumulator is

$$r_{c1} = r_{c0} + \Delta r_c, \tag{6}$$

where r_{c1} is the value of the compensating rate this time and r_{c0} is the value of the compensating rate last time.

Once receiving the synchronization frame, the clock synchronization unit compensates the frequency of clock counter.

(c) Clock Drift Averaging. The clock drift averaging calculates the mean of the latest eight times clock drift to store in the register DIFFTIME. The master reads the register to determine whether the process of synchronization is completed.

3.2.4. Microcontroller (μC_If) Interface. The μC interface (μC_If) is the connection between real-time MAC and micro-processor. The internal structure of μC_If is shown in Figure 10. The μC_if for CASNET has two 512 b RAM and registers. One RAM is Trans_buf ram used for transmit buffer and the other is Rev_buf ram used for receive buffer. The μC_If for microprocessor has 10-bit address, 16-bit data, Chip Select signal (cs), read/write enable and interrupt control signal.

3.2.5. Interrupt Control. An interrupt control selectively deals with the interrupt based on the priorities of four interrupt sources. Interrupt source supports edge trigger mode only.

Figure 11: System configuration for test.

4. Verification and Simulation of the Implemented CASNET

4.1. CASNET Slave Implementation. Using the Quartus II toolset, the present authors have implemented CASNET slave in a Cyclone II family FPGA, EP2C5, which offers high performance and high density. This version includes 4608 Les, 26 M4k Ram Blocks, 117 Embedded Memory, 13 18-bit × 18-bit Embedded Multipliers, 2 PLLs, and maximum 158 user I/O pins. The Quartus II synthesis and route summary indicate that the CASNET slave costs 3569 LE and 8.6 k ram block bits, and the max frequency is 78.2 MHz.

4.2. Test System Configuration. The testing environment, as shown in Figure 11, is composed of the CASNET master (CNC) and 50 CASNET slave boards, each of which is designed by using FPGA technology. The master station adopts the PC.

4.3. CASNET Cycle Time. Cycle time is one of the key variables of the motion control system. CASNET cycle time is made up of a weighted sum of master packet forwarding time, master and slave PHY delay, slave forwarding delay and propagation delays along the cables. In order to test the minimum cycle time, the first step is to calculate the minimum cycle times with a payload of 60 bytes per slave at a given constant time. As the second step, it is to calculate the minimum times that can successfully send and return a frame with a payload of 60 bytes per slave. The number of devices ranges from 1 to 50. The results are shown in Figure 12.

4.4. Time Synchronization Accuracy. Time synchronization accuracy indicates the maximum deviation between the slave clocks and the master clock. The oscilloscope CH1 connects the master device and the oscilloscope CH2 connects the slave device. The phase between the master clock and the slave clocks is shown in Figure 13. The time synchronization accuracy is 15 ns. The time synchronization accuracy is between 15 ns and 100 ns. The number of devices ranges from 1 to 50. The minimum deviation between the salve clocks and the master clock is 0~15 ns. The time synchronization

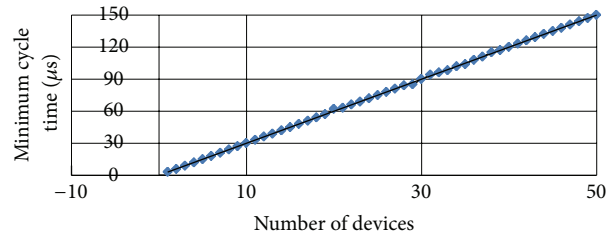


Figure 12: Minimum achievable cylcle times on a line topology network with 60 bytes payload per devices.

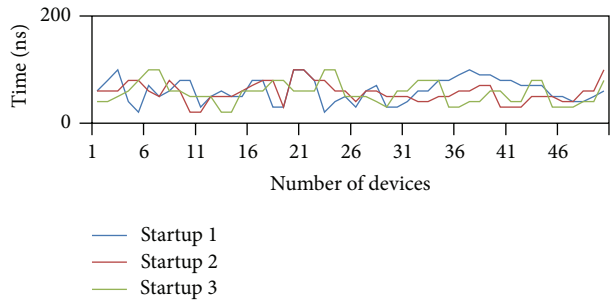


Figure 13: Synchronization accuracy at different startups.

accuracy is achieved at different startups. The average time from startup to synchronization is 20 s.

4.5. Wireshark Testing. Wireshark is a network packet analyzer that can capture live network traffic or read data from a file and translate the data to be presented in a format that the user can understand. Wireshark can define user packet to analyze. The authors of this paper use a PC as a master device. Two ways have been adopted to test packet loss rate. One is that the slave generates a random data with a certain payload to send to PC and writes the data to a file every 2 ms. The files are input to the Wireshark and compare the data with the data captured by the Wireshark at the master device. After the experiment with 50 slaves lasted for as long as 48 h, there is no error at all. The other is that the master device generates a packet and then sends the packet to the slaves. The slaves insert data to the packet and then return the packet to the master. Wireshark generates the data and captures the return data to analyze. No error occurs during long-time testing. These tests have proved that the packet loss rate is very low.

5. Conclusion

The present paper proposes a real-time Ethernet named CASNET. Instead of utilizing Ethernet MAC, the MAC is modified to meet the real-time requirements and is implemented in Verilog HDL. In other words, the MAC can provide deterministic behavior by processing the functions in hardware such as synchronization of a frame, address recognition, and CRC comparison. The CASNET is tested in terms of its cycle time, synchronization accuracy, and reliability. This solution has been proved feasible and the next step is to optimize hardware block and fabricate the

application specific IC (ASIC) for the CASNET. Certainly, it is necessary to carry out a complete process of functional verification through a field test with the fabricated ASIC.

Acknowledgments

This work is supported by National Major S&T Program of China (Grant no. 2009ZX04013), National Hi-tech Research, and Development Program of China (863 Program, Grant no. 2011AA04A104).

References

[1] J.-D. Decotignie, "Ethernet-based real-time and industrial communications," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1102–1117, 2005.

[2] M. Felser, "Real-time ethernet—industry prospective," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1118–1129, 2005.

[3] P. Neumann, "Communication in industrial automation—what is going on?" *Control Engineering Practice*, vol. 15, no. 11, pp. 1332–1347, 2007.

[4] S. Vitturi, L. Peretti, L. Seno, M. Zigliotto, and C. Zunino, "Real-time Ethernet networks for motion control," *Computer Standards & Interfaces*, vol. 33, no. 5, pp. 465–476, 2011.

[5] Standard for a precision clock synchronization protocol for networked measurement and control systems, IEEE 1588, 2008.

[6] IEC. IEC 65C/358/NP. Real-time Ethernet: SERCOS III, 2004.

[7] IEC. IEC 65C/355/NP. Real-time Ethernet: ETHERCAT, 2004.

[8] IEC. IEC 65C/356/NP. Real-time Ethernet: POWERLINK, 2004.

[9] IEC. IEC 65C/359/NP. Real-time Ethernet: PROFINET IO. Application layer service definition & application Layer protocol specification, 2004.

[10] Digital data communications for measurement and control-Fieldbus for use in industrial control systems, IEC 61158, 2003.

[11] Digital data communications for measurement and control—part2: additional profiles for ISO/IEC 8802-3 Based Communication Networks in real time applications, IEC 61784, November 2007.

[12] Z. Shao, C. Xue, Q. Zhuge, M. Qiu, B. Xiao, and E. H.-M. Sha, "Security protection and checking for embedded system integration against buffer overflow attacks via hardware/software," *IEEE Transactions on Computers*, vol. 55, no. 4, pp. 443–453, 2006.

[13] Y. Wang, D. Liu, Z. Qin, and Z. Shao, "Optimally removing Inter-Core communication overhead for streaming applications on MPSoCs," *IEEE Transactions on Computers*, vol. 62, no. 2, pp. 336–350, 2013.

[14] J. P. Thomesse, "A review of the fieldbuses," *Annual Reviews in Control*, vol. 22, pp. 35–45, 1998.

[15] H. Zimmermann, "OSI Reference Model-The ISO model of architecture for open system interconnection," *IEEE Transactions on Communications Systems*, vol. 28, no. 4, pp. 425–432, 1980.

[16] M. I. Heywood and A. N. Zincir-Heywood, "Register based genetic programming on FPGA computing platforms, in Genetic programming," in *Proceedings of the EuroGP*, R. Poli, W. Banzhaf, W. B. Langdon, J. F. Miller, P. Nordin, and T. C. Fogarty, Eds., vol. 1802 of *Lecture Notes in Computer Science*, pp. 44–59, Springer, Berlin, Germany, 2000.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 598647, 10 pages
http://dx.doi.org/10.1155/2013/598647

Research Article

Self-Assembling for Swarm Modular Robots Using MIMO Fuzzy Control

Tianmiao Wang,[1] Haiyuan Li,[1] and Cai Meng[2]

[1] *School of Mechanical Engineering and Automation, Beihang University, Xueyuan 37, Beijing 100191, China*
[2] *School of Astronautics, Beihang University, Xueyuan 37, Beijing 100191, China*

Correspondence should be addressed to Haiyuan Li; haiyuanvip@163.com

Received 19 April 2013; Accepted 12 June 2013

Academic Editor: Shao Zili

Modular robot are said to construct a diversity of morphogenesis with self-assembling strategies. They bring about an adaptive entity to deal with complex tasks. By analyzing integration design in module, perception, and control in detail, a swarm modular robot is presented with self-assembly scenario. Then, these active docking robots use the distance measured by infrared sensors between itself and edge of assembled structural entity as input. We design the fuzzy sets and if-then rules according to the human reasoning in following process. Based on Mamdani-type inference, the fuzzy controller can yield two outputs. The outputs are, respectively, used as steering angle speed and linear speed. Due to the diversity of self-assembled structure by the swarm modular robots, we conduct a series of simulated experiments. The results demonstrate the effectiveness and efficiency of the proposed controller in swarm robots' edge-following process.

1. Introduction

Modular robots have been viewed as a new approach to rapidly manufacturing kinds of structural objects and some symbiotic robotic configuration using autonomous development methodologies [1–3]. In this subarea recently attracting the robotic communities, self-assembling of a swarm of robots is widely developed as an efficient mechanism enabling shaping what is desired [4–6]. In spite of physical or intelligent limitation (such as size, computing, power, and manipulation) of independent incapable robots in certain circumstances, the swarm robots can use self-assembly method to form inter-robot physical connection to generate a stronger robotic entity.

A swarm of modular robots in a self-assembly system generally is homogeneous, and some rich resources including hardware and software are embedded in each one by mechatronics. Compared with traditional fix-structure robots, the self-assembly system incorporating modularity offers significant flexibility and economic advantage by changing the connectivity between different robotic modules or parts. In some extreme conditions such as extravehicular activity (EVA) on orbiting spacecraft or planetary surface, these modular systems can be employed at any time to assemble into manipulators or coexplorer space robotics to replace or assist the crew. One scenario can be envisioned that, before the crews reach Mars, a large number of robots have assembled into Mars houses including furniture [7].

According to the functional or geometric characteristics, there is a large body of research works, respectively, focused on different types of modular robotics. Distributed swarm robots are equipped with the appropriate grasping or connecting mechanisms to assemble into physically coupled morphologies [8–12]. These robots use low-level control logic to determine the connection sequences and distributed controller to connect with the base or beacon. With group-level cooperation, the connected swarm robots can make collective responses, for example, cooperative transportation or moving on rough terrain. Each robot is self-mobile and capable to interact with others, but the resulting entity is incapable to stand up and be extended in 3D space.

Without considering the engineering limitation, one subarea of morphogenetic robotics [13] focuses on the

cognitive development, which concentrates the codesign of the controller and morphogenesis [14]. Based on lattice-type self-reconfigurable robots, a few evolutionary or bioinspired algorithms are employed to generate artificial structures or biological lifeforms [15, 16]. From potential engineering viewpoint, collective constructions combine the passive blocks and active mobile robots [17]. Inspired by social insects, the mobile ground robots [18, 19], aerial vehicles [20, 21], or external propulsion/fabrication device [2, 22, 23] can collect the modular passive blocks to construct complex structure which usually is stationary and however not given the capacity of whole body's locomotion.

Combining the advantages of the robots mentioned above, swarm modular robots in distributed mode are able to disassemble or assemble into different morphogenesis which either are functional structures or locomotive organisms [24–26]. These swarm modular robots often move on the ground and can self-assemble into 2D planar organism, and afterward, can reconfigure into 3D structure. These self-mobile and interactive robots are more flexible to construct the composite entity without external direction or assist. However, the design of the higher degree of modularity integrating components and embedded system is required as well as adaptive algorithms [27]. In our previous works, a swarm modular robot is proposed. These independent robots self-assemble into different configurations, and self-assembly based on finite state machine [28] and whole body locomotion [29] are verified. Different from our previous work, in this paper, we focus on self-assembling process of one single robot. Considering successful development of fuzzy control on robot navigation and following [30–33], a many-input many-output (MIMO) fuzzy control method is introduced to enhance self-assembling function according to an analysis of the robot navigation in stage of seeking the assembled entity's open docking direction. The proposed method improves the efficiency of self-assembling.

This paper is organized as follows. In Section 2, we describe the swarm modular robot system including structure, perception, and function and give some illustration of their constructing entity. In Section 3, the experiments and general control method based on finite state machine (FSM) are presented. In Section 4, we introduce the fuzzy controller in self-assembling process in detail. This controller allows the active robot can follow the partially or fully assembled entity. In Section 5, we implement the experiments where a robot followed a diversity of morphologies steadily. Finally, we conclude and discuss future work.

2. Swarm Modular Robot

The swarm modular robot with self-assembly function is designed to satisfy the capabilities of autonomous morphogenesis from a number of elementary components. It is featured that each modular robot to operate and move as a single self-contained individual with the capacity of self-propulsion, perception, computation, and power supply. The second significant capacity of the modular robot is to aggregate into different morphogenesis and disaggregate any time through docking elements and information sharing. Besides, evolution capacity incorporating sensing, computing, and responding in a changing environment is a key issue for autonomous morphogenesis.

2.1. Robot Module. This study was conducted on the groups of swarm modular robotic system, which consists of a number of self-propelled modular robots called *Sambot* (see Figure 1). The *Sambot* robot synthesizes characteristics of swarm robots and self-reconfigurable robots. Each *Sambot* is a wheeled mobile robot which is composed of a cubic main body and an active docking interface. There is a pair of docking hooks on active docking interface whilst there is a pair of docking grooves on each of four lateral interfaces. On the two wheels on the bottom of the main body, the robot can use the active docking interface to grasp any passive docking interface on four laterals of the other robot's main body. Thereby, a swarm of *Sambots* can autonomously move and self-assemble into certain kinds of morphological structures by physical connections with each other. When two robots are connected, the active one can rotate the other through rotating joint.

2.2. Perception. Each robot is surrounded by a group of infrared (IR) sensors. Each pair of sensors consists of an emitter and a receiver, shown in Figure 1. They are divided into three classes according to their functionalities.

(i) IR detecting sensors. These sensors are installed on the upside of the front of the main body, and they can measure the remote range and detect certain height of objects and other robots, as shown in Figure 1.

(ii) IR approximating sensors. On each lateral side of the main body, there are two pairs of sensors to measure the distance between the robot and others, as shown in Figure 2(a).

(iii) IR docking sensors. Four pairs of sensors are mounted on the active docking interface for docking alignment or work as "IR approximating sensors" in self-assembly stage, as shown in Figure 2(b).

On one hand, these IR sensors could measure the distance by emitting signals and receiving them reflected from another object. On the other hand, they can receive the signals from other robots' emitters in order to align the active docking interface and the passive one.

2.3. Inter-Robot Freedom and Their Morphogenesis. From degree of freedom view point, the modular robot features four degrees of freedom (dof). Each robot can move on the two differential derived wheels, having three degrees of freedom on the plane (see Figure 3(a)). An additional dof is introduced to allow the active docking interface rotating around the main body (see Figure 3(a)). According to docking direction of the two connected robots, two connected robots can fix the in between dof, including two robots' rotation axes orientation parallel (see Figure 3(b)) or orthogonal (see Figure 3(c)).

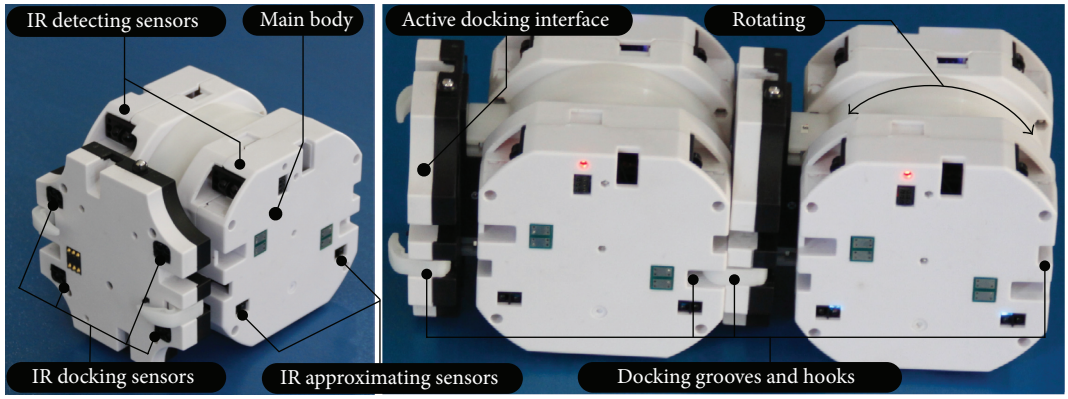


Figure 1: Swarm modular robot (*Sambot*). Size: 102 mm length, 80 mm width, 80 mm height; weight: 350 g, including: battery (2 hours), ZigBee communication, and MCU.

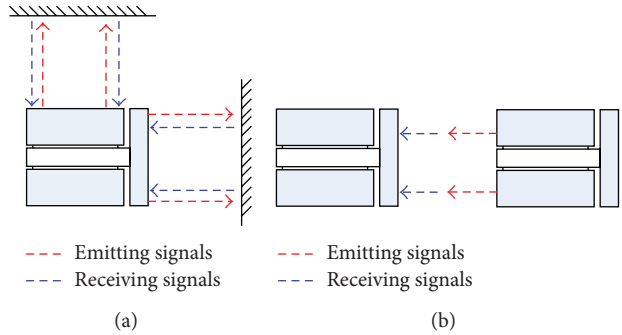


Figure 2: (a) Measuring the distance by emitting signals and receiving them reflected from another object. (b) Receiving the signals from other robots' emitters.

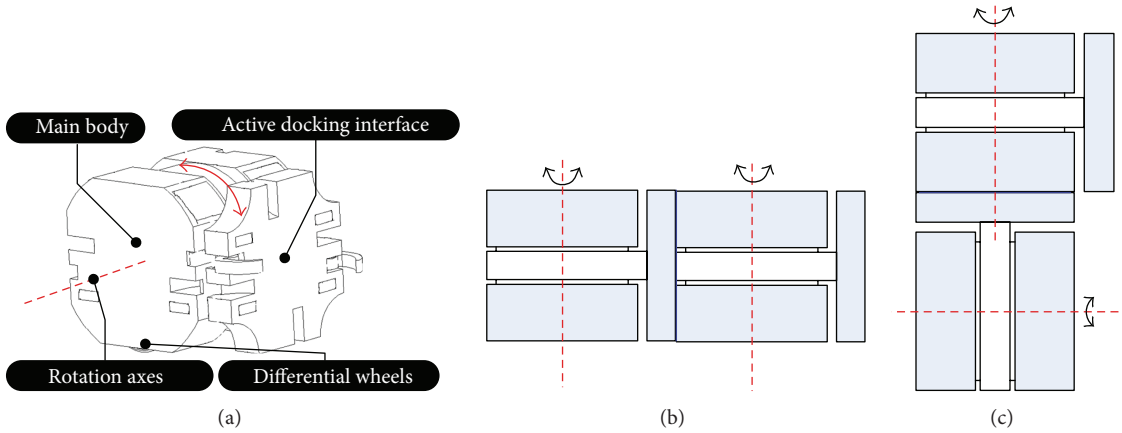


Figure 3: Actuation generating degree of freedom and orientation of the rotation axes. (a) Rotation axes orientation. (b) Parallel axes. (c) Orthogonal axes.

Simultaneous connection to a robot is available, and thereby these modular robots can construct different configurations of organisms (see Figure 4) through rearrange of degree of freedom. These morphogeneses are brought about according to high-level command or adaptive self-assembly.

2.4. Software Architecture. For achieving distributed control and modularity in software as well as hardware [34, 35], the different layers of software development are shown in Figure 5. On the bottom, some basic modular functional components such as sensors, direct servo, and MCU are

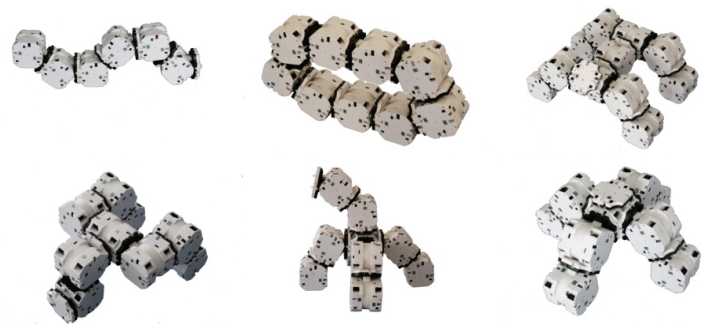


Figure 4: Diverse morphogenesis including the locomotive organism and manipulating tool.

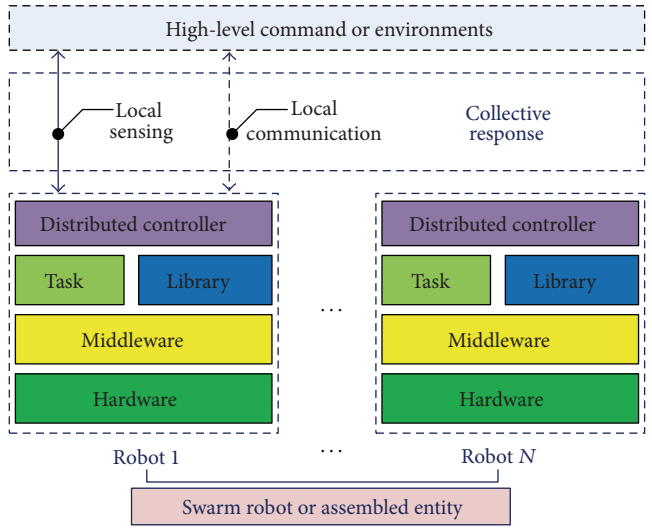


Figure 5: Software architecture embedded in robotic system. The programs in each robot are identical and distributed to adapt to the task requirements.

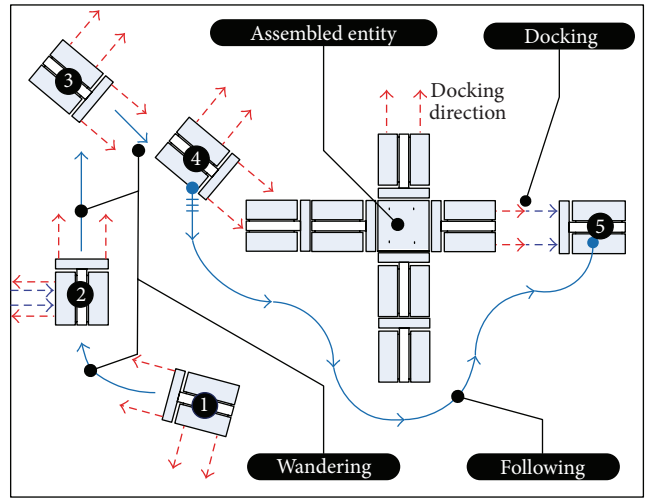


Figure 6: Self-assembly process based on finite state machine in the sided experimental arena.

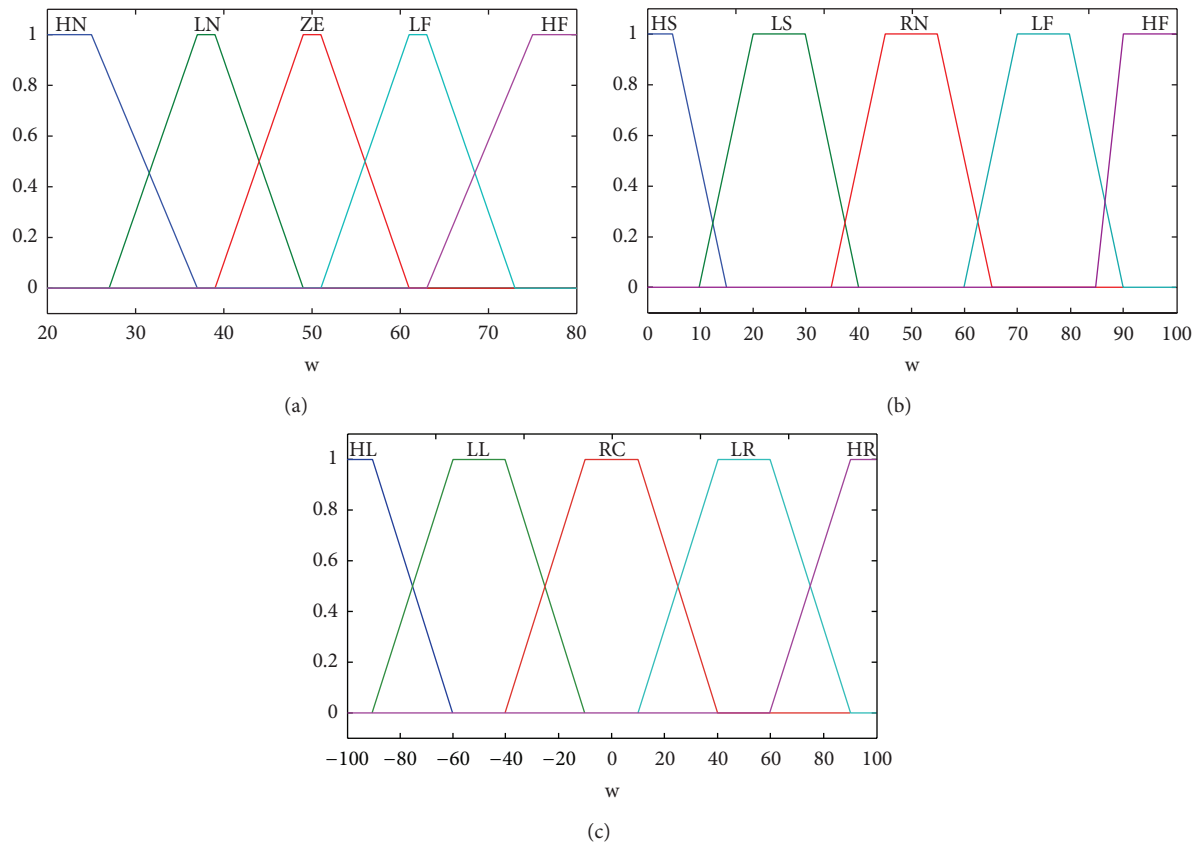


FIGURE 7: Input and output fuzzy sets and the corresponding, membership functions. (a) Input: d_{LF}, d_{LB}, d_{FL}, and d_{FR}. The distance respectively, measured by the front left, front right, left front and left back sensors are located five sectors which partially overlap. The obtained four input variables have the same distribution function in discourse domain. (b) Output: v. Five overlapping sectors describe the intensity of the linear speed from stop to maximum. (c) Output: ω. To avoid head-on or lateral collision to an edge, the steering orientation is divided into left and right.

standardized to open to the upper layer. Middleware is on one way used to define the unified interfaces, on the other way to maintain local communication with close or vicinal robots using CAN bus when connected or Zigbee when disconnected. A large number of perception or motion libraries are integrated into the MCU, and some energy and fault management tasks are predesigned to load when they are necessary. Each robot has two states. When swarm robots work without connection, they use distributed controller including behavior-based or target-oriented subcontroller to make a collective response to the environments or high-level command. When they are assembled, they are either commanded by the center or make decision by distributed controller.

3. Experiment Scenario and Self-Assembly Strategy

A swarm of robots is distributed inside the platform, a table surrounded by a fence. They search for the partially assembled entity or a robot who recruits the surrounding robots. In our previous work [28], a self-assembly method based on the finite state machine is proposed to grow the structure. The process of self-assembly has three main stages: wandering, following, and docking (including locking).

First, one robot, called SEED, broadcasts signals over the air to attract other robots and simultaneously opens its IR sensors on the docking interface as "docking" directions. Then, other separate robots begin to wander for searching for the docking directions. In wandering stage, as shown in Figure 6, the robot search for assembled entity using the combination of IR detecting sensors and IR approximating sensors. If it meets the fence before finding assembled entity (see ① to ② in Figure 6), it would follow along the fence (see ② to ③ in Figure 6) until the corner and switch to approximate the assembled entity along the diagonal of this corner (see ③ to ④ in Figure 6); if otherwise, it would detect the assembled entity and directly switch to following stage. Then, the active robot will follow the edges of assembled entity until it captures the docking directions using its IR approximating sensors installed on the left side (see ④ to ⑤ in Figure 6). When one of IR signals representing docking direction is detected by the robot, the robot would align itself with the passive docking interface on the assembled entity and dock with it (see ⑤ in Figure 6). After that, the active docking robot becomes a part of the assembled entity and opens IR sensors to be docked by

Table 1: Fuzzy controller including the input and output.

On front interface	Left sensors →	Fuzzification	Fuzzy inference	Defuzzification	→ linear speed v
	Right sensors →				
On left interface	Front sensors →				→ steering speed ω
	Back sensors →				

Table 2: Fuzzy rules for edge-following using d_{LF}, d_{LB}, d_{FL} and d_{FR} as input and yielding v and ω as outputs. In the process, d_{FL}, and d_{FR} measured by the two front sensors are large, indicating no objects in the front of the robot.

(a) Rules for obtaining v to keep the appropriate linear speed

d_{LB} \ d_{LF} (d_{FL} is HF and d_{FR} is HF)	HN	LN	ZE	LF	HF
HN	LS	LS	LS	HS	HS
LN	HS	LS	LS	HS	HS
ZE	HS	LS	HF	LS	HS
LF	HS	HS	LS	LS	HS
HF	HS	HS	LS	LS	LS

(b) Rules for obtaining ω to keep in parallel to the edge

d_{LB} \ d_{LF} (d_{FL} is HF and d_{FR} is HF)	HN	LN	ZE	LF	HF
HN	HR	HL	HL	HL	HL
LN	HR	LR	LL	HL	HL
ZE	HR	HR	RC	HL	HL
LF	HR	HR	LR	LL	HL
HF	HR	HR	HR	HR	HL

Table 3: Fuzzy rules for obstacle avoidance using d_{FL} and d_{FR} as input and yielding v and ω as outputs. In the process, d_{LF} and d_{LB} measured by the two left sensors are not taken into consideration.

(a) Rules for obtaining v to keep the appropriate linear speed

d_{FR} \ d_{FL}	HN	LN	ZE	LF	HF
HN	HS	HS	HS	HS	HS
LN	HS	HS	HS	HS	HS
ZE	HS	HS	HS	HS	HS
LF	HS	HS	HS	HS	HS
HF	HS	HS	LS	LS	×

(b) Rules for obtaining ω for avoidance and lateral alignment

d_{FR} \ d_{FL}	HN	LN	ZE	LF	HF
HN	HR	HR	HR	HR	HR
LN	HR	HR	HR	HR	HR
ZE	HR	HR	LR	HR	HR
LF	HR	HR	LR	HR	HR
HF	HR	HR	LR	LR	×

others according to the topology. In previous work, a method incorporating connection state table and tree-like graph is proposed to conduct self-assembly. Once a robot docks with the assembled entity, it can lookup connection state table and determine its position in graph to get connection state in order to decide which sensors to open.

Considering the improvement of the self-assembly algorithm, we analyze the three stages. Efficiency of wandering stage depends on the experimental arena, and docking stage costs less time than other two stages. Thereby, we focus on the following stage and use fuzzy control to conduct edge-following. Through edge-following, the individual robot can navigate smoothly along the edges of the assembled entity for seeking docking directions.

4. Fuzzy Controller for Edge-Following in Self-Assembling

Due to the diversity of morphology which can be assembled by a swarm of robots, the shapes of the edge are changing. The fuzzy controller based on Mamdani model is designed to guide the robot to follow the edges of the assembled entity.

4.1. Inputs, Outputs, and Fuzzification. Infrared sensors measure the distance between the lateral side of the robot and the opposite edge of the assembled structure. As shown in Figure 2(a), the left front, left back, front left, and front right distances are, respectively measured and represented by the linguistic variables d_{LF}, d_{LB}, d_{FL}, and d_{FR} as input variables of the fuzzy controller. Each input variable is divided into five fuzzy sets: hard near (HN), little near (LN), zero (ZE), little far (LF), and hard far (HF).

The outputs of fuzzy control are linguistic variables v and ω, respectively, representing the moving linear speed and steering speed of the mobile docking robot. v is described by five fuzzy sets: hard slow (HS), little slow (LS), run normal (RN), little fast (LF), and hard fast (HF), and ω is described by three fuzzy sets: hard left (HL), little left (LL), run center (RC), little right (LR), hard right (HR).

Thereby, the belongingness of each input or output to a particular fuzzy set is determined via the trapezoidal membership function (MF) as follows:

$$\mu(x) = \begin{cases} 0 & : \quad x < a \\ \left(\dfrac{x-a}{b-a}\right) & : \quad a \le x < b \\ 1 & : \quad b \le x < c \\ \left(\dfrac{x-c}{d-c}\right) & : \quad c \le x < d \\ 0 & : \quad x > d. \end{cases} \quad \text{or} \quad (0 \ : \ a = b) \quad \text{or} \quad (0 \ : \ c = d) \tag{1}$$

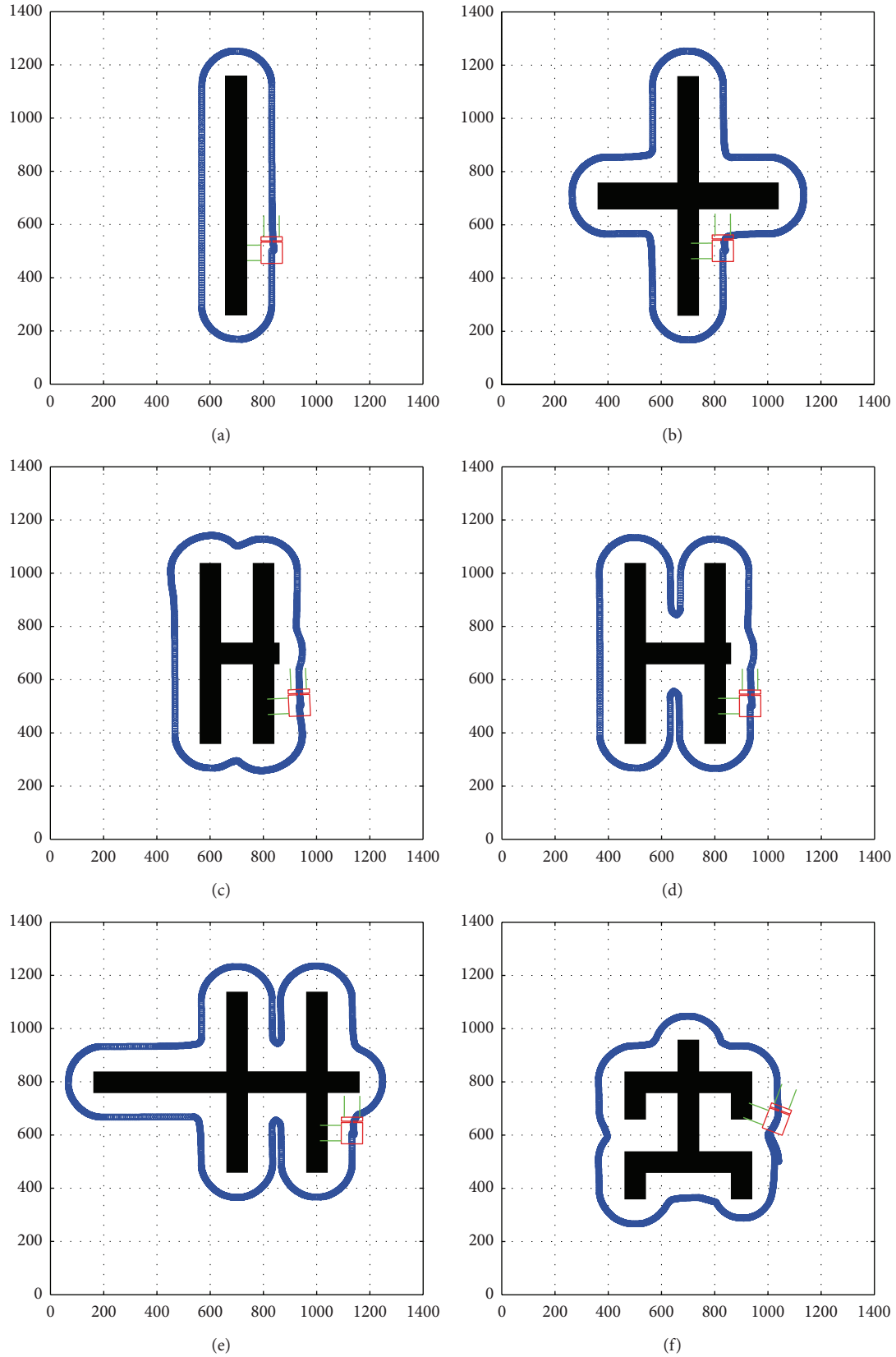


Figure 8: Morphogenesis consisting of a number of robots and trajectory when the active docking robot follows the edges. These structures can be assembled in 2D surface and followed. (a) Snake-like configuration. (b) Octopus-like configuration. (c) Little quadruped configuration. (d) Large quadruped configuration. (e) Lizard-like configuration. (f) Human-like configuration.

Table 4: Records of the edge-following process shown in Figure 8. These factors provide evidence for analysis of controllers.

Configuration	Snake	Octopus	Quadruped (little)	Quadruped (large)	Lizard	Human
Number of robots	9	15	15	16	17	22
Time	20	40	32	37	55	45
Number of concavo corners	0	4	0	4	4	6
Number of convex corners	4	8	4	8	12	12

According to the measuring distance by IR sensors and movement capacity of the robot and safety margin, the universe of discourse of the fuzzy set of inputs d_{LF}, d_{LB}, d_{FL},and d_{FR} is [20 80] mm. Considering the maximum speed, the universe of discourse of the fuzzy set of outputs v and ω is, respectively, [0 100] mm/s and [−100 100] degree/s. The fuzzy controller is described in Table 1, and membership functions of input and output are illustrated in Figure 7.

4.2. Fuzzy Rule Base. In the edge-following process, the rule base for both wall following and obstacle avoidance is constructed based on the human reasoning. The following three metarules are employed to design fuzzy rule base.

(i) If the robot moves forward at an appropriate distance from edge of assembled entity without any other structure in the front, the robot steers straight.

(ii) If the robot deflects to the left, heading against the assembled entity, or deflects to the right, going away, then the robot would turn towards the opposite direction to correct a deviation.

(iii) If there exit parts of assembled in the front of the robot, the robot would turn right to align the left side of the robot with the front edge.

Here, we divide the fuzzy rule base into two subbases, one of which is used to achieve edge-following and the other for avoidance and alignment. According to the above metarules, we use d_{LF}, d_{LB}, d_{FL}, and d_{FR} as inputs and, respectively, yield v (see Tables 2(a) and 3(a)) and ω (see Tables 2(b) and 3(b)) as outputs. In edge-following stage, we assume these exist no parts of the assembled entity at the front of the robot (d_{FL} is HF and d_{FR} is HF) and only use the d_{LF} and d_{LB} as variables, as shown in Table 2. Regarding the avoidance and alignment stage, without consideration of d_{LF} and d_{LB}, only d_{FL} and d_{FR} are adopted, as shown in Table 3.

4.3. Fuzzy Inference and Defuzzification. We use a Mamdani-type fuzzy inference to formulate the mapping from the sensed distance to control action. Based on fuzzy rules mentioned above, the outputs of each rule are combined into a single fuzzy set, and we use the center of area (COA) defuzzification method to resolve two output values, respectively obtaining robot linear speed v and the steering angle ω.

5. Experimental Simulation and Discussion

According to the morphology of assembled entity, we design a number of structures to verify the fuzzy controller. In the following process (mentioned in Section 3), the active robot would follow the edge of partially assembled entity to seek the docking direction which might occur anywhere and towards any possible direction. In this section, considering the fulfillment of structural shapes in different situations, we conduct the simulation in MATLAB to make the robot follow the whole edges.

According to a diversity of morphogenesis mentioned in Section 2, we use the swarm modular robot to assemble some structural configurations on planar surface, and an active robot is placed at a distance from the assembled entity. In Figure 8, the assembled structures consisting of a large number of identical robots are simplified in black. The active robot will follow the edges by detecting the black pixels. These structures mimic some shapes of animals. In Figures 8(a) and 8(b), a snake-like structure composed of nine robots and one composed of thirteen robots are connected in black block representing the modular robot, the active robot followed the boundary smoothly. In Figures 8(c) and 8(f), there are two short sections of concave boundary. In Figures 8(d) and 8(e), two long sections of concave boundary lead the active robot into it, and the active robot still can escape from the conflict and succeed in following.

Table 4 lists the number of the robots constituting the black entity and shows the time of the robot's following whole edges. Regarding these facts that influence efficiency of following, number of concave corners cause the robot to take more time to adjust itself. In the previous work [28], when it is expected to dock with snake-like and quadruped structure of five robots, it took up to 100 s to follow the edges to search for the docking direction based on state-based behavior method. Obviously, fuzzy controller in this study is an improvement.

6. Conclusions

The distributed swarm modular robot can move itself to explore which robots claimed for docking. Through a self-assembling method, groups of these robots can aggregate into certain structures beyond the individual limitation in body and brain. In this study, we in detail present the capacity of swarm modular robots including the hardware, perception, and software, which can address elements of autonomous construction. A fuzzy controller is applied in the self-assembly process for seeking the docking direction. For a large number of structural entities, we implemented the

controller and analyzed the efficiency. It is obvious to shorten the time to follow the edges of the assembled entity.

In the future, we will implement distributed control in swarm robots and self-assembled structure and build an online response strategies using self-assembly to enhance adaption to the environments. Considering respective advantages of the evolutionary computing and fuzzy control, we expect to further improve fuzzy controller on the inefficient or ineffective aspects in the self-assembly process.

Acknowledgments

This work is supported by the National High Technology Research and Development Program of China ("863" Program) (2012AA041402), National Natural Science Foundation of China (Grant nos. 61175079 and 51105012), and Fundamental Research Funds for the Central Universities (Grant no. YWF-11-02-215).

References

[1] P. Moubarak and P. Ben-Tzvi, "Modular and reconfigurable mobile robotics," *Robotics and Autonomous Systems*, vol. 60, pp. 1648–1663, 2012.

[2] K. Gilpin and D. Rus, "A distributed algorithm for 2D shape duplication with smart pebble robots," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '12)*, pp. 3285–3292, 2012.

[3] Y. Terada and S. Murata, "Automatic assembly system for a large-scale modular structure - Hardware design of module and assembler robot," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '04)*, pp. 2349–2355, October 2004.

[4] P. F. Damasceno, M. Engel, and S. C. Glotzer, "Predictive self-assembly of polyhedra into complex structures," *Science*, vol. 337, pp. 453–457, 2012.

[5] R. Groß and M. Dorigo, "Self-assembly at the macroscopic scale," *Proceedings of the IEEE*, vol. 96, no. 9, pp. 1490–1508, 2008.

[6] G. M. Whitesides and B. Grzybowski, "Self-assembly at all scales," *Science*, vol. 295, no. 5564, pp. 2418–2421, 2002.

[7] A. Spröwitz, S. Pouya, S. Bonardi et al., "Roombots: reconfigurable robots for adaptive furniture," *IEEE Computational Intelligence Magazine*, vol. 5, no. 3, pp. 20–32, 2010.

[8] L. Murray, J. Timmis, and A. Tyrrell, "Self-reconfigurable Modular e-pucks," in *Swarm Intelligence*, pp. 133–144, Springer, Berlin, Germany, 2012.

[9] R. O'Grady, R. Groß, A. L. Christensen, and M. Dorigo, "Self-assembly strategies in a group of autonomous mobile robots," *Autonomous Robots*, vol. 28, no. 4, pp. 439–455, 2010.

[10] W. C. Evans, G. Mermoud, and A. Martinoli, "Comparing and modeling distributed control strategies for miniature self-assembling robots," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '10)*, pp. 1438–1445, May 2010.

[11] R. Groß, M. Bonani, F. Mondada, and M. Dorigo, "Autonomous self-assembly in swarm-bots," *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1115–1130, 2006.

[12] R. Groß and M. Dorigo, "Evolution of solitary and group transport behaviors for autonomous robots capable of self-assembling," *Adaptive Behavior*, vol. 16, no. 5, pp. 285–305, 2008.

[13] R. Doursat, H. Sayama, and O. Michel, "Morphogenetic engineering: reconciling self-organization and architecture," in *Morphogenetic Engineering*, pp. 1–24, Springer, Berlin, Germany, 2012.

[14] Y. Jin and Y. Meng, "Morphogenetic robotics: an emerging new field in developmental robotics," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 41, no. 2, pp. 145–160, 2011.

[15] R. Thenius, M. Dauschan, T. Schmickl, and K. Crailsheim, "Regenerative abilities in modular robots using virtual embryogenesis," in *Adaptive and Intelligent Systems*, pp. 227–237, Springer, Berlin, Germany, 2011.

[16] Y. Meng, Y. Zhang, and Y. Jin, "Autonomous self-reconfiguration of modular robots by evolving a hierarchical mechanochemical model," *IEEE Computational Intelligence Magazine*, vol. 6, no. 1, pp. 43–54, 2011.

[17] K. Petersen, R. Nagpal, and J. Werfel, "Termes: an autonomous robotic system for three-dimensional collective construction," in *Proceedings of the 7th Robotics: Science and Systems*, 2011.

[18] S. Magnenat, R. Philippsen, and F. Mondada, "Autonomous construction using scarce resources in unknown environments," *Autonomous Robots*, vol. 33, pp. 467–485, 2012.

[19] J. Werfel and R. Nagpal, "Extended stigmergy in collective construction," *IEEE Intelligent Systems*, vol. 21, no. 2, pp. 20–28, 2006.

[20] Q. Lindsey, D. Mellinger, and V. Kumar, "Construction with quadrotor teams," *Autonomous Robots*, vol. 33, pp. 323–336, 2012.

[21] J. Werfel and R. Nagpal, "Three-dimensional construction with mobile robots and modular blocks," *International Journal of Robotics Research*, vol. 27, no. 3-4, pp. 463–479, 2008.

[22] G. Mermoud, M. Mastrangeli, U. Upadhyay, and A. Martinoli, "Real-time automated modeling and control of self-assembling systems," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '12)*, pp. 4266–4273, 2012.

[23] K. Gilpin, K. Koyanagi, and D. Rus, "Making self-disassembling objects with multiple components in the robot pebbles system," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA '11)*, pp. 3614–3621, 2011.

[24] S. Kernbach, B. Girault, and O. Kernbach, "On self-optimized self-assembling of heterogeneous multi-robot organisms," *Studies in Computational Intelligence*, vol. 355, pp. 123–141, 2011.

[25] W. G. Liu and A. F. Winfield, "Autonomous morphogenesis in self-assembling robots using IR-based sensing and local communications," in *Swarm Intelligence*, pp. 107–118, Springer, Berlin, Germany, 2010.

[26] S. Kernbach, E. Meister, F. Schlachter et al., "Symbiotic robot organisms: Replicator and Symbrion projects," in *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems (PerMIS '08)*, pp. 62–69, August 2008.

[27] H. Wei, Y. Chen, J. Tan, and T. Wang, "Sambot: a self-assembly modular robot system," *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 4, pp. 745–757, 2011.

[28] H. Wei, H. Li, J. Tan, and T. Wang, "Self-assembly control and experiments in swarm modular robots," *Science China Technological Sciences*, vol. 55, pp. 1118–1131, 2012.

[29] H. Wei, Y. Chen, M. Liu, Y. Cai, and T. Wang, "Swarm robots: from self-assembly to locomotion," *The Computer Journal*, vol. 54, no. 9, pp. 1465–1474, 2011.

[30] F. Cupertino, V. Giordano, D. Naso, and L. Delfine, "Fuzzy control of a mobile robot," *IEEE Robotics and Automation Magazine*, vol. 13, no. 4, pp. 74–81, 2006.

[31] S. F. Desouky and H. M. Schwartz, "Genetic based fuzzy logic controller for a wall-following mobile robot," in *Proceedings of American Control Conference (ACC '09)*, pp. 3555–3560, June 2009.

[32] H. Boubertakh, M. Tadjine, and P.-Y. Glorennec, "A new mobile robot navigation method using fuzzy logic and a modified Q-learning algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 21, no. 1-2, pp. 113–119, 2010.

[33] M. A. K. Jaradat, M. H. Garibeh, and E. A. Feilat, "Autonomous mobile robot dynamic motion planning using hybrid fuzzy potential field," *Soft Computing*, vol. 16, no. 1, pp. 153–164, 2012.

[34] H. Wei, B. Wang, Y. Wang, Z. Shao, and K. C. C. Chan, "Staying-alive path planning with energy optimization for mobile robots," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3559–3571, 2012.

[35] Y. Wang, D. Liu, Z. W. Qin, and Z. L. Shao, "Optimally removing inter-core communication overhead for streaming applications on MPSoCs," *IEEE Transactions on Computers*, vol. 62, no. 2, pp. 336–350, 2013.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 969304, 15 pages
http://dx.doi.org/10.1155/2013/969304

Research Article

A General Framework of Mechatronic Modular Architecture

Ying Zou, Gang Zhao, and Tianmiao Wang

School of Mechanical Engineering and Automation, Beihang University, Xueyuan Road No. 37, Haidian District, Beijing 100191, China

Correspondence should be addressed to Ying Zou; yy506@126.com

Received 10 April 2013; Accepted 22 May 2013

Academic Editor: Hongxing Wei

The independent designs guided by domain technologies of mechatronic system lead to the problem of hard compatibility and the difficulty to reuse resources between the systems. This paper presents an approach based on the reuse for modular mechatronic system, and a general framework for modular mechatronic system is designed. After the analysis of related characteristics of the mechatronic system, the modular principle of mechatronic system is proposed. Secondly, a general framework (F-Model) based on reusability of software and hardware in mechatronic system is developed. Following is to elaborate in details on the parts of software, hardware, and infrastructure of mechatronic system. Finally, 6-DOF industrial robot arm and medical accompany robot are taken as our experiment platform to design system based on F-Model. Results in practical application confirm the value of F-Model for reusing modules in designing mechatronic systems.

1. Introduction

Depending on the abundant application technology in mechatronic system (MS) developed over the past decades, such as motion control, path planning, position, trajectory tracing, visual processing, and trajectory interpolation, various kinds of mechatronic systems have been developed. Design approaches for MSs guided by the domain technologies above presented are independent and hard to be compatible mutually, so the problems faced by the people in the fields of study and development in the processing of designing MSs are that how to apply those technologies more easier, how to simplify design process, how to lower design difficulty, how to reduce repetitive work, and so forth. In order to have more time for researchers to study further more on the basis of previous work, it is very important to collect the valuable study achievements contributed by previous researchers for freeing them from repetitive work. The key problem in the MS design is to reuse software and hardware.

Monolithic programming is a common method adopted for software design in traditional MSs. The diversity of application domain of MSs would like to create very different architectures on software and hardware. So it is hardly realized not only to use general approach to design MSs and to reuse software and hardware but also to guarantee quality of MSs.

To solve these problems, researchers and institutes have started to research the reuse approaches of mechatronic system, which focus on some aspects as below.

Shield the differences from different hardware using universal network middleware, such as RT-Middleware [1, 2] and MIRO [3]. Due to the huge cost of the resource and the diversity of the hardware in MSs, it is difficult to transplant universal network middleware on the different platforms. Universal middleware taken up a large number of resources may have an effect on the implementation of RT-task in system with less resources and low processing speed.

Researchers and institutes have proposed some hierarchical architecture models of robot software system, such as K-Middleware [4] and ORiN [5], which are used to solve the heterogeneity between the modules building up MS. But these models are aimed to shield heterogeneity from hardware level, and the generalized reuse model is still not established now.

MS software architectures based on Component-Based Software Engineering (CBSE) are gradually accepted while some theories, such as OROCOS [6], RT-Middleware [7],

MSRS [8], emphasize the reuse of partial software models, but the structural reuse method is not concerned when the hardware platforms have changed.

On the basis of the theory of software architecture and CBSE, Davide Brugali summarized the current study achievements and presented challenges faced by the design to robot software, called 4Cs [9, 10]: Computation, Configuration, Communication, and Coordination.

Additionally, some studies solved partial problems on compatibility while the software architectures based on component were not adopted and the reuse model of software and hardware in MS was not considered at all. These studies include Open-R [11], uPnP [12], ASEBA [13, 14], and CLAWER [15].

In fact to construct the hardware-independent software for MSs is really quite difficult. Though the researches above have considered the differences among different hardware levels, the differences are so great that the meaningful reuse of MS resources is not realized between the hardware. The MS controllers can be simple as MCU or be complex as industrial server, and thus their corresponding software system can hardly be the same. Designing the software taking MS features themselves into no account can cause that the reuse difficulties are concentrated to platform-independent software framework. This is not suitable for the MSs which usually have more changing hardware and less systematic resources.

Based on CBSE approach we have presented robot modular architecture based on functional component [16] and its software framework [17, 18]. In this paper, the general framework for modular MS is summarized based on software and hardware separation approaches and the MS model characteristics. The framework is expressed by the view of software and hardware reuse.

The principle is presented in this paper for MS modularity reuse and the general framework for MS is designed, and, based on the framework, two kinds of mechatronic instances are constructed. Firstly, the characteristics of mechatronics are analyzed and the modularity principles are presented. Secondly, a model and its framework are presented based on software and hardware reuse in mechatronics modular architecture. Thirdly, the software part, hardware part, and systematic infrastructure of the framework are discussed. Finally, based on the framework, two kinds of system instances, 6-DOF industrial robot and medical robot, are implemented.

2. Modularity Analysis of Mechatronic System

2.1. Modular MS Model Analysis. Traditional MS hardware module includes mechanical part and electronic part. The mechanical part decides the system physical motion, while the electronic part decides the information flow. The relationship of the two parts lies in the following: the motor transforms the information into the physical motion of the motor axis, which realizes the transformation from the information world to the physical world. The sensor transforms these quantities such as the sound, light, force, and heat in physical world into the information data, which realizes the backtracking from the physical world to the information world. The physical world is complicated and various, and the information representation can be unified by computer information processing. As a specification aiming to share and communicate information, the software and hardware reuse of MS is typical information subject of Domain-Specific Software Architecture (DSSA), and should be based on the connotation of information transfer and data flow.

Take MS computer control for instance, the core control part of MS mainly contains digital computer, signal amplifier, sampling holder, and so forth, and it mainly includes the digital information output to motor and sampling from the sensors; the controlled part mainly includes mechanical bodies, transmission devices, motors, and sensors, and through the information transforming, the controlled subsystem expresses the physical motions and state changes in overcoming the external force, self-gravity, kinematic pair damping, and so forth. Additionally, the generalized controlled object also includes the external environments where the MS is located. The environments are influenced by the controlled subsystem and re-act to it. There is a classical control loop between the control system and the controlled object (Figure 1).

For the computing capability is enhancing, traditional control system is gradually replaced by computer control system, and control functions are implemented by computer software procedures and run on the way of miniaturization, integration, and embedded application. The computer system has the standardized I/O feature, and thus the transfer function modules in traditional control model are packaged to these computer I/O modules, such as driver output module and the sampling holder module, while the control algorithms and other functional requirements are designed in computer software. Hence the modularity of MS control part is essentially the modularity of system software and electronic parts.

The research [19] shows that connecting the advantages of the top-to-bottom design approach based on the domain model and the component design approach is effective in DSSA design. The mechanical part and electronic part of MS reflect different views of domain (Figure 2) and have their respective modular model approach. Based on the function separation and the open system requirement MS software model should not couple the kinematics model and control model. For once some coupling produced, its design ideas would be backtracked to the traditional design road which is led by some domain professional technology, and thus the reuse capability and compatibility of the systematic resources would be decreased. In the other side, the MS software model limited by systematic resources cannot be the same with the universal computer software model but should be the simplification or new construction in some dimensions.

2.2. MS Modular Software and Hardware Separation Principle. In the architecture based design the software usually follows the CBSE development specifications, encapsulates every reusable software module using component, and supplies

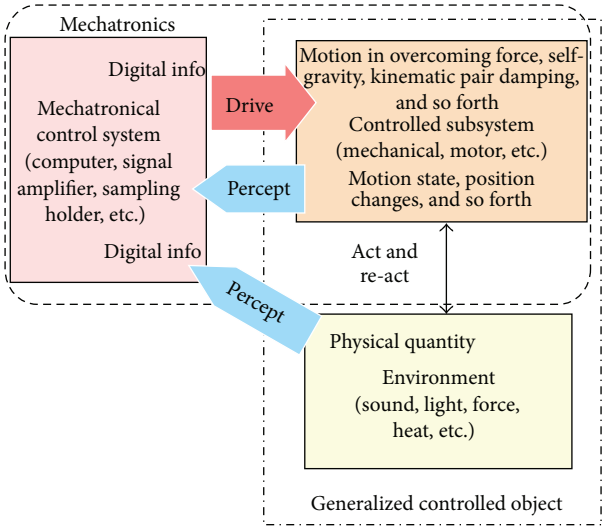


Figure 1: Control model of MS.

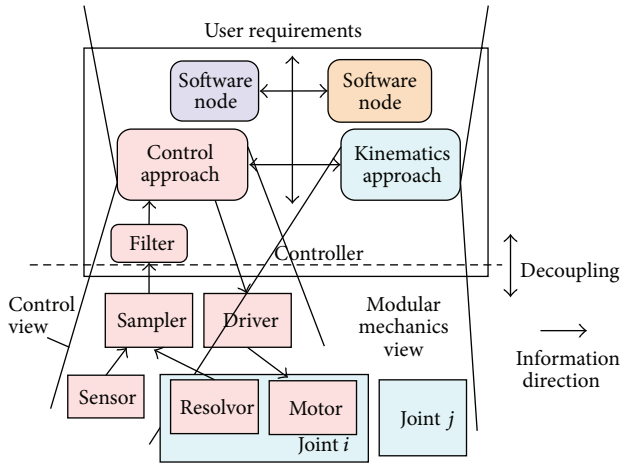


Figure 2: Different views of modular mechatronic control system.

standard mechanism for visiting and operating. However, for it is quite hard to be compatible with different MS hardware platforms, if the component based run-time framework is designed without considering the hardware level, the software scale will be very large, and the difficulties from software reuse will be concentrated to the framework and then it will be quite difficult to design a platform-independent run-time framework. Plenty of researches above have discussed this hardware and software compromise and concluded that there are difficulties in running the traditional software middleware in these MSs.

The origin problem in MS reuse is the various software and hardware, and hence the core point for modular reuse is decoupling the MS software and hardware. To encapsulate the domain feature module and to define the standard interfaces for the communicating, are the reuse conditions of MS modules.

Therefore, the MS software and hardware modularity should follow the basic principles as below.

(a) Structure Separation Principle (Software and Hardware Separation). Considering the MS design's labor division and efficiency, the structure separation of hardware and software should be implemented. As the MS has strong specificity, the holistic design based on the domain technology is applied. Structure separation can concentrate the domain technologies and decrease the coupling degree between software and hardware, and then the displacement between different modules can be implemented.

(b) Function Separation Principle (General and Special Separation). Function separation refers to the separation of function connotation which the MS module represents. For the function is the external characteristic expressed repeatedly by module or unit, the function separation must be based on the determination and stability of module function characteristics.

In MS the function bound is usually determined. Therefore, these functions can be extracted and abstracted and furthermore the abstracted results can be classified into universal function and special function based on the domain characteristics. The function stability decides the stability of architecture style, thus it is suitable to apply clear architecture pattern in MS.

(c) Combination Principle (Classification). MS function has the features of combine and cluster; for example, some servo motors have the position detection function, and some manipulators can combine the perception sensor function. Therefore, function modules of modular MS should have the combined capability based on the granularity expansion. And the functional component modules should have the similar interfaces and install rules with the original single component.

(d) Open Principle (Extension). About the MS structure, in order to improve the reuse capability, source code opening, resources sharing, and the function resource extension support are critical. In fact, to open the system design is also are important method to improve the quality of software, and many researchers have already researched the connotation of openness.

The principles above settle the main points of this paper.

2.3. Granularity Choice. Granularity is the opposite conception of coupling degree, and it is a relative conception but not an absolute conception. Granularity is data unit refinement or comprehensive degree level. The higher it is refined, the smaller granularity is; inversely the lower it is refined, the larger granularity is. The data granularity is a design subject; and granularity choice in different level will influence the information interaction or combined efficiency between the data modules [20].

In the MS modular design it is suitable to choose the large granularity in order to make some flexibility for module design and interface install. In other words, as the software and hardware are various, it is not suitable to describe the MS by tiny granularity and large-scaled and detailed specifications. Additionally, if the tiny granularity is chosen for the new hardware device, the component combined form in the system design will be complex and special, and the designers will be trapped into massive detailed design constraints, which will greatly increase the design difficulties and workloads.

Take the modular separation of a mobile robot system, for example, (Figure 3), to depict the system by *Sonar Sensor*, and so forth, is small granularity design, while to depict the system by *Sensor Group*, and so forth, is large granularity design. It is seen that the large granularity approach simplifies the system design work and decreases the difficulties.

2.4. Network Structure of Modules. There are different comprehensions in academic researches between the modular MS conceptual models. The researches on module reuse problem mainly focus on the modular modeling of mechanical modules, modularity of software system, modularity of control system, and so forth. However, the modular modeling problem of swarm-based or cell-based robotics is not the decoupling and reusing problem of the module; thus this paper does not discuss this issue.

In domain engineering the modular modeling researches of mechanical part and control part have gotten some progression, for example, modelling of modular MS and building approach based on spinor [21, 22], modular reconstruction or reconfiguration of control system [23, 24], and so forth. These domain engineering researches solved the modularity problem within the domain, thus when considering the information dataflow and the module reuse problem in the design, these models can be encapsulated in modules and meanwhile fully play the domain role. For example, the mechanical module problem can be transformed into information problem as motor control module, motion control module, kinematics software module, and so forth, which can accommodate the modular model of mechanical motion.

From the view of information model, traditional control structure can be expressed as different groups formed by subdivision network node (Figure 4). The MS control structure has three types: centralized structure, distributed structure, and hybrid structure. As the physical structures are different, system software and hardware resources are distributed in the network.

In Figure 4 *Network Node* expresses distributed application. In the computer control system, the essence of these three structures is the centralization and distribution of control application procedures, and thus the control function application is organized as an analysable network structure between the MS modules. In this way, the complex network connection model is expressed and the connection can be handled in the MS control.

The network node connections form the logical topology relationship. However, it is not the same as the topology from the physical connections. The implementation of data interaction between the modules has no relation with the topology itself, because the network communication protocols usually support multiple topologies. In MS the logical topology expresses the control structure, and in the software modules of control algorithms or simulation model, it is the important configurable information for system structure.

The topology independent characteristic of network protocol determines that wherever a device module in the network located, it can be visited from allopatry limitedly. In fact, the network topology formed by the network nodes is always limitedly "transparent" in front of the device modules, and hence the device module in the whole system should be managed in an unified way, in order to ensure that every application procedure can finally limitedly visit the device modules via the network.

2.5. Brief Summary. From the analysis above, a brief summary can be given for modular MS framework.

(a) MS modular architecture model should not couple with modular kinematics model and modular control model and should independently design and encapsulate these domain functions as domain modules.

(b) Although the decoupling of software and hardware adds system resource cost to a certain degree, it will improve the flexibility and maintenance of the system design. The adding resource cost cannot be too much, for the MS resource limit mentioned above. The decoupling is the necessary method to realize the module reuse and to choose the mature modules can greatly improve the quality of the design.

(c) The large granularity is chosen for the scale of modules and the architecture framework should have loose connecting relationship and more compatibilities. The detailed processing should be settled into the modules, and the necessary communication channels should be reserved between the modules.

(d) Considering characteristics and resources of the runtime platform, the traditional software architecture should be clipped, simplified, and modified necessarily. Within this process, it is important to specially consider the design of the MS concurrent, real-time, and priority capabilities and implement the corresponding structures and infrastructures.

3. General Framework

3.1. F-Model Conception. Based on the MS characteristics analyzed above, the modular architecture model, F-Model, is presented to construct kinds of MSs. Differing from other approaches to solve software reuse problem, F-Model is constructed in MS information model to reuse systematic resources with the approach of hardware and software separation and interface standardization. F-Model system mainly includes the software structures by CBSE theory, hardware structures based on functional classifications, and the systematic infrastructures to provide standard visiting

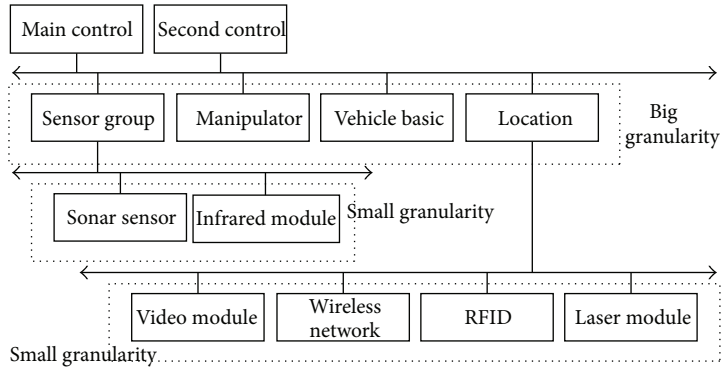


Figure 3: Modular separation instance of a mobile robot.

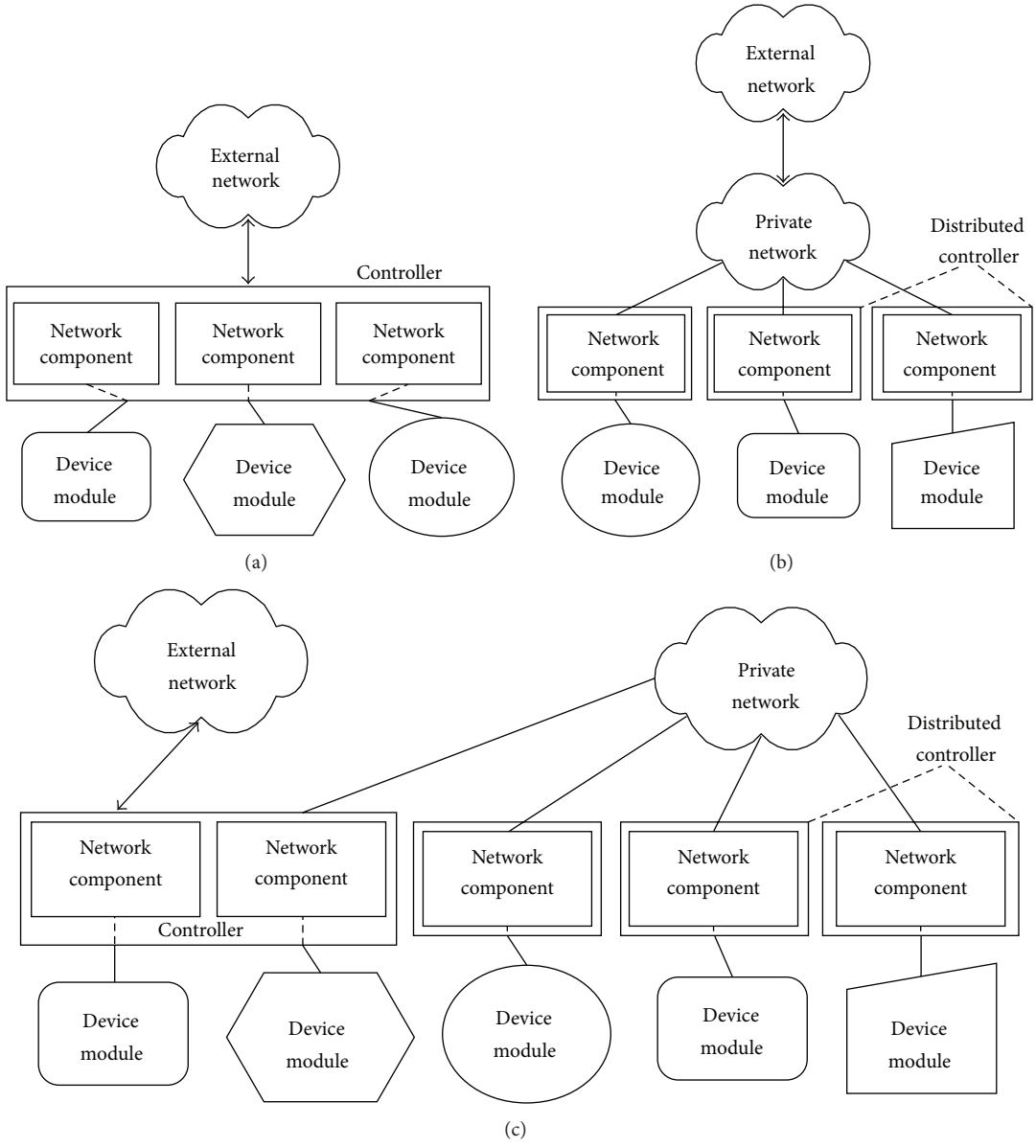


Figure 4: Network structures of modules.

interfaces. When designing system with reusable resources, software components and abstracted hardware modules can be invoked by application procedures in different robots.

F-Model can be expressed as hierarchy model with robot application layer, service layer, component layer, middleware layer, device driver layer, and hardware layer (Figure 5).

(1) *Application Layer* realizes MSs' control and basic interactive application by associating and organizing function components.

(2) *Service Layer* is to provide some basic services for application, such as management service, error handling, and power management.

(3) *Component Layer* includes basic components and programs that are necessary to modular MS, and they are materialized in the upper layer through the abstraction of special function.

(4) *Middleware Layer* includes a middleware core that can continually monitor dynamical process of components, find the abstraction form up to down, and shield the difference in device layer.

(5) *Device Layer* mainly provides hardware device driver program compatible with operating system (OS) and device protocol, and so forth.

(6) *Hardware Layer* defines the correspondence for MS electrical modules and expresses the transformation from physical quantity to information data quantity in system.

3.2. Componentization and Separation of Software and Hardware in F-Model. In the architecture defined by F-Model, the separation of software and hardware applies the abstraction for decoupling. Following the information path, the system hardware part is abstracted and classified by the module kinds, while the software is decoupled by resource functions and characteristics. The module interfaces are encapsulated. A unified infrastructure and framework are applied to load the modules and a manager module is constructed to manage the hardware and software modules.

That is, in F-Model, the reusable granules include architecture, applications, components, manage procedures, and hardware device modules, while the granules, such as universal data package, common protocol, and common objects, will be treated as basic system resources and will not be encapsulated specially in order to keep the flexible developing room. The design considering of F-Model is to reserve the MS core control model and then to separate the hardware and software; thus the problem is solved that the reuse difficulties are concentrated to software framework, which is caused by the hardware differences.

3.3. Modular MS Based on F-Model. The construction process of modular system is the process of module assembling. It is supported by a resource manager, software components, hardware abstract classes, the real-time services provided by the real-time system scheduling, and so forth, and finally a series of application is released, as shown in Figure 6.

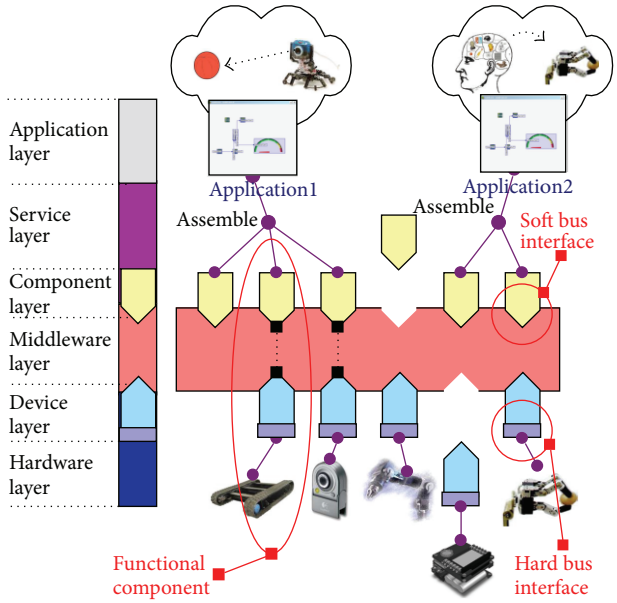


Figure 5: Architecture and hierarchy model.

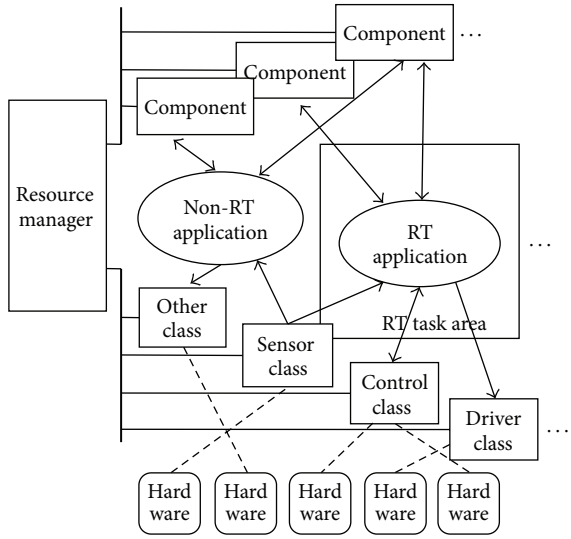


Figure 6: Modular system by F-Model.

The control structure of modular MS is settled; that is, the core domain model is considered. Based on this, the MS is decomposed by the functions and the independent applications to be implemented are divided into domain application and universal applications. As the application is decoupled with the hardware, it is easy to know that these applications can be reused in different platforms.

Based on the system requirement the respective priorities are settled for application procedures and the real-time tasks and non-real-time tasks are chosen. Domain applications usually have high requirements of running priority and resources.

Design the application, respectively, and then, the design considering is how to choose the relative components and

hardware classes and the interface timing and mode between them, and how to handle the data coordinating and sharing within the process.

Finally, the system is assembled in the component cabin based on the framework and the final system is released.

4. Software Modularity Structures

4.1. Software Module Conception. In CBSE system component provides settled services and follows the Producer-Consumer pattern, and the consumers need not concern the implementation details of the components. In MS, as the hardware modules in lower layer are usually domain-dependent, the component after separating will be domain-dependent, while the applications for users in upper layer are usually non-domain-dependent. Therefore we separate the applications and the resource components in F-Model, and set the decoupling in the component layer to domain and not domain. And then a holistic resource manager is designed to manage the components. When assembling the MS based on Figure 6, the different components can be chosen to implement the Model-Based-Design (MBD) domain applications and Component-Based-Design (CBD) non-domain applications (Figure 7 and the Control Domain frame expresses domain application). This design joins the advantages of MBD and CBD. Thus these components follow the advantages of CBSE design in reuse and decrease the difficulties of application design, and the quality can be guaranteed in the domain characteristics design.

For example, the MS control algorithm is complex, and some algorithms are not the simple combination in time or space with software and hardware modules. Encapsulating the algorithm as a single domain module can be used without influencing other modules.

Therefore the systematic component resources include multiple types such as the classes, monolithic procedures, and library procedures; to use such settings it is more flexible and more effective than to encapsulate the software to classes. This is critical in the MS platforms which have limited resources. To be simple, in following sections we will not specially discuss the domain component.

4.2. Real-Time Support for Application Node. The separation of application nodes is good for real-time and non-real-time distributions, and the MS real-time application is definite domain designing which can be formally analyzed. Based on the definition of real-time capability and influenced factors, the real-time task is critical link for the real-time application requirement. Within the full data flow, as long as the real-time capability of every link is guaranteed, the real-time capability of the whole control logic chain is guaranteed. Typical data flow is as shown in Figure 8.

In Figure 8, the rectangles express the components, the ovals express the buffer data structures in the producer-consumer process, and there are no direct invocations between the producer and the consumer. The fillet rectangles express the relative hardware modules. The directed thick lines express the systematic real-time data path. The directed

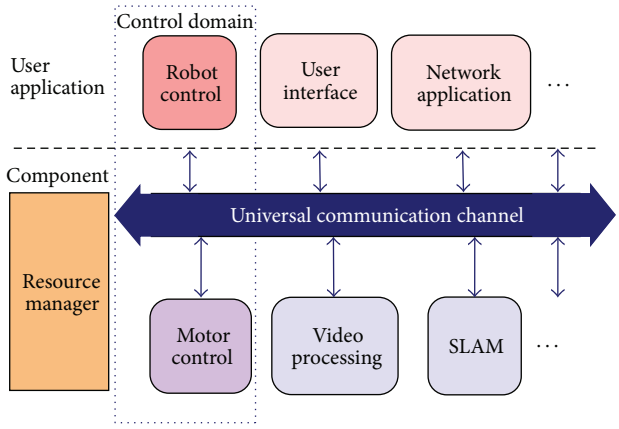


Figure 7: Software modularity structure.

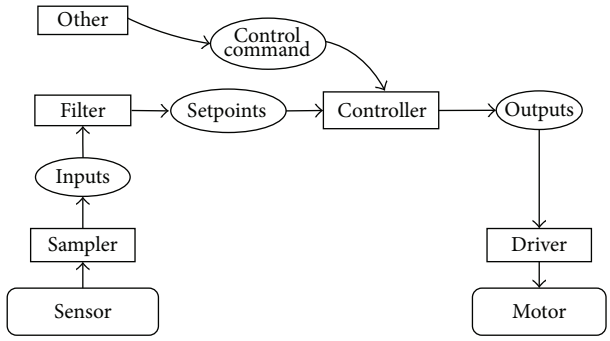


Figure 8: Dataflow instance of real-time node.

graph formed by the components and connectors is the typical Pipe-Filter pattern, and the sequential structure between components is formed. There are no loop locks in the graph; thus the real-time capability of flow executing results can be ensured by this structure.

By the model analysis above of MS information model, the transfer process from the *Sampler* via the *Controller* to the *Driver* is necessary in MS domain model, and then the real-time chains in the graph will be formed. Each link of the chains can be distributed in multiple real-time tasks, even though the real-time network is distributed in real-time tasks in different network platforms. The advantage which lies in this structure is platform-independent, even as the single loop chip such as MCUs, the efficient real-time application and distributed application can be constructed.

4.3. Distributed Structure of Application Node. F-Model provides the data chain structure of real-time guarantee and provides the real-time capability by the RTOS task scheduling. The multiple OSs supplying tasks in different levels can simultaneously run in the system.

Suppose there are the real-time application node A and the non-real-time application node B, and based on the running platform kinds and the network distribution situation, the node allocations of A and B have several kinds as follows: A and B run on one RTOS on one platform; A and B run on different OS and different platforms with network

connection; A and B run on different OS of one multicore platform; A and B run on the hybrid structure of the above three.

(a) A and B Run on One RTOS. Node A and node B create the real-time tasks using the task creating method of RTOS (Controller B in Figure 9). It is known from the real-time research that the efficiency will be decreased when real-time task of B runs.

(b) A and B Run on Different OSs of Network Platform. Node A and node B run on the different hardware platforms and via the real-time network connection (Controllers B and C in Figure 9). The several RTOS and non-GPOS, respectively, run on different platforms in network and nodes A and B are implemented by respective system tasks.

(c) A and B Run on Different OSs of the Same Platform. Many MSs usually have single computer or controller platform and the MS nodes need to run on the same platform. This means that GPOS and RTOS should simultaneously run on the platform. The common method is adding real-time modifications such as RTAI [25], to construct a new real-time core to meet the requirement.

We apply the hybrid OS framework RGMP [26] to provide the integrating real-time services on the multicore platform. The clients of nodes are ported to the real-time OS to provide the distributed real-time services.

RGMP is a software framework that can run in multiple OS and it is mainly for multicore processors. Different OSs run on the different cores of one multicore processor, and the real-time task and nonreal-time task are supplied by different cores using RGMP; the illustration is in Figure 9 Controller A.

(d) Hybrid Structure. In the hybrid structure (Figure 9 Controllers A, B, and C) the RGMP, RTOS, and GPOS can be used simultaneously for different controllers.

It is seen that by the RGMP and the F-Model real-time framework, the complex structure of lower layer is abstracted and standardized into an operation framework that can provide the real-time and non-real-time tasks. The corresponding parts coupled by the OS and task properties are separated and the platform-dependency is lowered.

5. Hardware Modularity Structures

5.1. Basic Principle. By the principle of F-Model, the abstraction and classification of hardware modules should be constructed in MS information model.

In Figure 10, the hardware device modules have multiple electrical interfaces, and the modules are loaded to operation system with corresponding device driver procedures; the unified device operation interfaces are allocated by the OS, the device open, read and write, configure, reset, and so forth, and these device operations are implemented. The resource manager reads the registry and configuration information

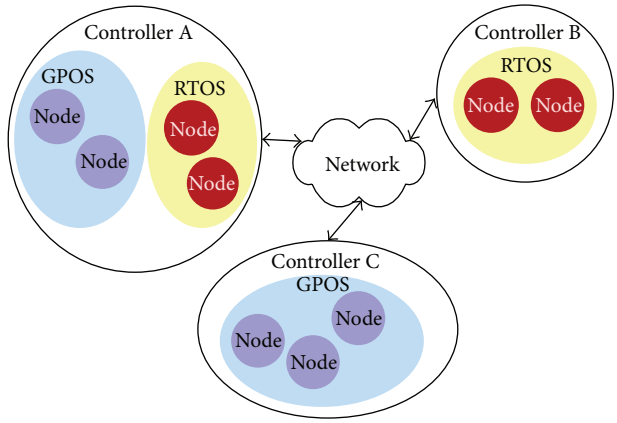


Figure 9: Distributed network and real-time capability.

from the devices, and then classifies them to corresponding hardware class.

The application in the upper layer can visit the resource manager to look up and enumerate the classes encapsulated by the hardware modules if necessary, and then the operating interfaces will be obtained. These classes can be reflected to the user state in the OS and show their feature interfaces by deriving. The application operates the hardware by the classes and after the system compiled, the high executing efficiency can be ensured. Encapsulating the hardware as class will lower the efficiency in a certain degree, however, the reusing and inserting of devices will be possible the device reuse is possible and it is easy for the inserting of the new devices.

The resource manager is an independent software module that can be transported between different OSs, and it is not coupled with device manager in OS; thus applications in different MS can visit the hardware module resources in the same classification.

5.2. Module Classification. Based on the MS information model the hardware modules are simply classified and these classifications are, respectively, encapsulated as the systematic universal components that are supplied to the real-time framework.

Driver Module (Information Output): Common MS driver module includes amplifiers, servos, and modular joints. The corresponding component is *Driver Class*.

Sensor Module (Information Input): Common MS sensor module includes sensors, resolvers, encoders, and various position switches. The corresponding component is *Sampler Class*.

Control Module (Information Input and Output): Includes various controllers, control boards, and cards. The corresponding component is *Controller Class*.

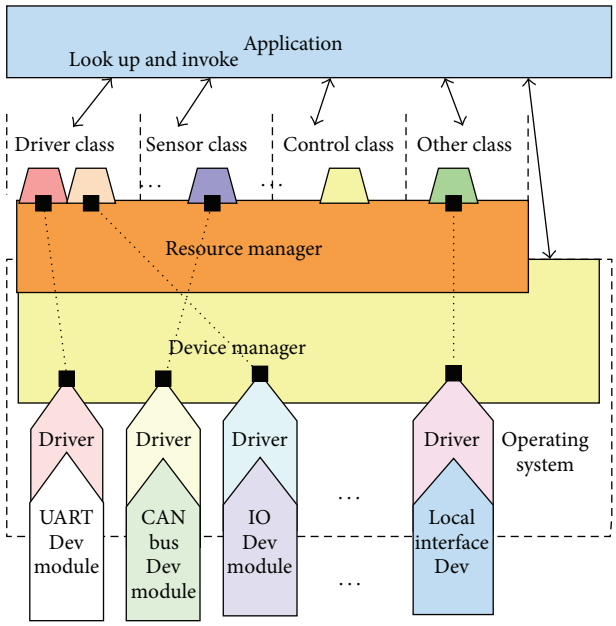


Figure 10: Hardware modularity structure.

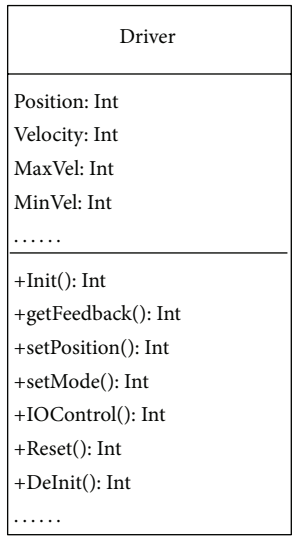


Figure 11: Driver class.

Another module includes external communicating module and HMI module. They also have their own corresponding components.

To be noticed, the module kind has totally different meaning from the electronic interface kind. The module output stands for the MS logical information output, and in output module can have electronic input data interface; for example, common amplifier has the bidirectional communication interfaces with the controller for configurations.

In Figure 11, we take Driver Class, for instance, and give its members and universal operating interfaces.

The hardware classifications express the abstraction of hardware modules and meanwhile they constrain the usage and configuration of the hardware resources, in order to implement their decoupling and reuse. Information, such as hardware identifiers, manufactures, hardware types, and electrical interface kinds is managed in the resource manager. The applications visit the resource manager for the necessary hardware modules.

The hardware is abstracted and encapsulated to classes, and this is the decoupled process between the device data and the application nodes and is also the decoupled process between the data and the model. In the OS, the existing devices are classified into stream device, block device, and so forth. However, the devices operate the hardware data without marking the device kinds in model. Devices in one classification may have different interface types; for example, there are motors with analog interface and motors with field bus interface, and hence if the classification based on the MS model is designed, the hardware decoupling can be implemented.

Also take the *Driver Class,* for instance, (Figure 12), in the application design, the output from the application node to the driver module is the velocity value of wheel rotation. It is a modular data which is not concerning what kind and what interface of motor is used in lower layer, and what data format is applied. The hardware abstract class encapsulates the model specifications and uses standard device interfaces of system in its methods, after deriving *Driver Class* the corresponding driver data operations can be implemented. If the new driver module is necessary, the corresponding driver procedures of OS should be added and then the abstract class is derived to implement the new interface operations. There is no need to change the application procedures and the decoupling is implemented.

To manage the hard modules the resource manager stores the property information and when a new module is inserted into the system, these information should be updated. The information format is as follows.

Identifier: 64 bits, includes 5 parts:

Class: 8 bits, to mark the module classification.

Function: 8 bits, to mark the module function.

Subclass: 8 bits, to mark the module's subclass.

Protocol: 8 bits, to mark the module protocol for electrical interfaces.

Des Code: 32 bits, is the manufacture information and defined by the manufactures.

5.3. Considerations for Hardware Module Compatibility. As the classes can be derived, the decoupled hardware modules can be in various changing, and then the devices can be inserted into system flexibly. The added device can be either local device or network device and the exact protocol should be included for compatibility.

As the MS controllers are various and their processing capabilities are different, essentially speaking, they are all computer units in different levels, and the network protocol levels to run on the controllers are also different. In the ample resource controller, such as industrial computers and

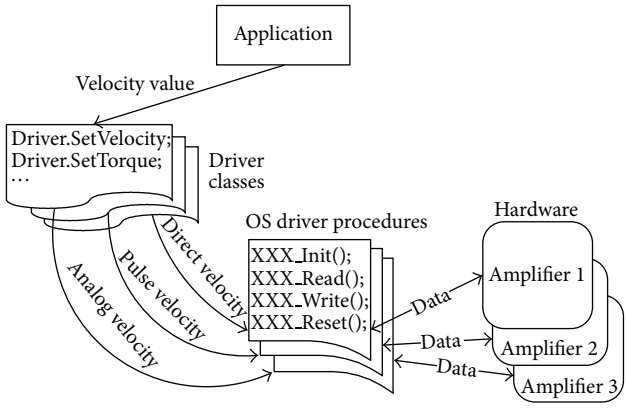


Figure 12: Hardware abstract class instance.

industrial servers, the full IP protocol or TCP/IP protocol can run well, while in the resource limited controller such as MCUs and DSPs, though the IP protocol stack can not be loaded, the proxy method [27] can be used and then the controller can be connected to the network in larger range by the identifier specification.

6. Infrastructures

6.1. Light-Weight Middleware. To the middleware in the framework, the support for different MS platform kind should be considered and the simple structures and better capability and system transporting capability should also be held. Common middleware is not suitable for the MS applications, for they are based on the TCP/IP and the middleware package is great with low efficiency. Based on researches of Magnenat et al. [13], the system cost of TCP/IP is usually more than Mbytes. Therefore, the common middleware has great difficulties applying in memory or storage resources limited systems. The middleware for MS should have the basic structure to load MS modules and should also have conveniently porting, clipping, and extending capabilities.

Additionally, common middleware is from the universal network middleware and its support for network is simple, and the support for common buses used in MS is not good. And the smooth route between the networks of multiple buses is hard to handle.

It is mentioned above that to apply a settled software middleware platform in various MS hardware will meet the problems of porting and supporting. And the MS middleware is suitable to apply the easily flexible and compatible structures and apply the simple interface in the modules' interactions. In this paper we use the identifier based interaction mechanism and standard client/server socket method based on IP protocol and then construct the communication between the modules.

6.2. Resource Manager. Resource manager is the important manage unit for modular MS, and it includes the resource manager and configuration function of software and hardware modules. For the visit rules is settled for the system modules each other, the resource manager can be considered as a simplified middleware module.

Resource manager includes the parts of resource manage core, component library, module address decoder, and parameter and configuration manager. It maintains the identifier of components and nodes and then the resources manage is implemented. Each identifier has a corresponding software component or hardware class in the component resource library. The user can visit these resources by the identifier without considering whether the component or node is local; in other words, resource manager redirects the visit operations to the real locations of the resources which are masked to applications.

New module needs to be registered to resource manager and then added to F-Model. Meanwhile the unique module identifier name will be confirmed and stored. Then all the application can search and visit the module by the identifier. After the application design is completed, the resource configuration of the current design can be saved to files by the resource manager.

6.3. Interaccession Mechanism between Modules Based on Identifier. Firstly, the component or node registers to the resource manager and asserts its identifier. Then the registered identifier is the unique identifier of components or nodes, and its identity can be identified by all components or nodes in the whole system. Identifier follows the subscribe-publish pattern; that is, one node gets another node's communication interfaces and parameters by subscribing the identifier and furthermore communicates with it using the provided service interface by settled protocol (Figure 13).

When the communication is established, the node can supply the settled services and standard socket for communications. The modules are decoupled and the simplified middleware mechanism by resource manager can meet the MS application's network requirements.

7. Implementations

7.1. Industrial Application—Welding Robot Arm

(a) Introduction of Welding Robot Arm. Modular Welding Robot (MWR) is standard 6-DOF joint robot (Figure 14). It can do settled operating task by human teaching and its end can track the lines, arcs, and free curves trajectories. With the welding machine and tools the MWR can repeat the welding workpiece task.

Applied hardware modules in MWR are the following: Modular Robot Controller Module, Modular Robot Joint Module, Sensor Modules, I/O Device Module, and so forth.

Applied software modules are the following: Robot Solution Module, Robot Type Configuration Module, Decoding Module, Rough Interpolation Module, Fine Interpolation Module, Axis Control Module, Debug and Error Handling Module, Human-Machine-Interaction Module, Logical Control Module, and so forth (Figure 15).

Other distributed modules: Teaching Module, Simulation Module, Remote Monitoring Module, and so forth.

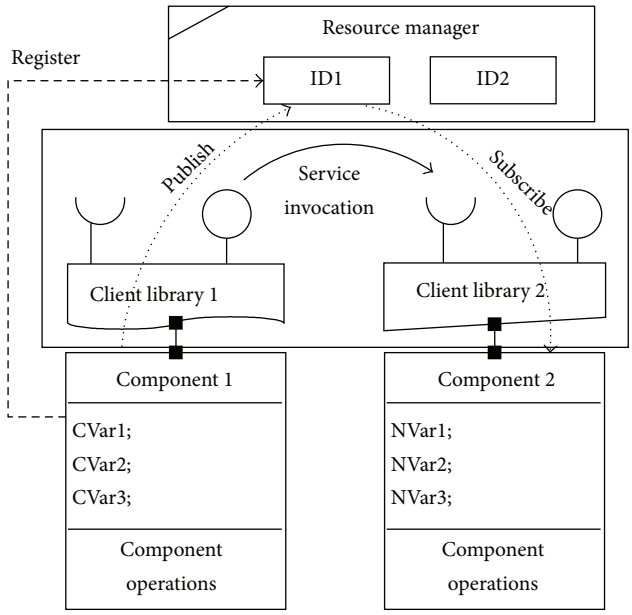


Figure 13: Communication approach between modules.

Figure 14: Modular Welding Robot (MWR).

The system resources are the following:

Processor: Marwell PXA270A 520 MHz

Memory: 128 MB

Storage: 128 MB Flash

Operating System: uCOS II.

(b) Structure of MWR. As common industrial 6-DOF joint robot system is not open, considering the structure characteristics of 6-DOF joint robot, the general method is to bind the 6-DOF joint robot body as a single part. Its servo drivers and motors are packaged and locked within the entity and they are hard to separate since the control and solution of robot are

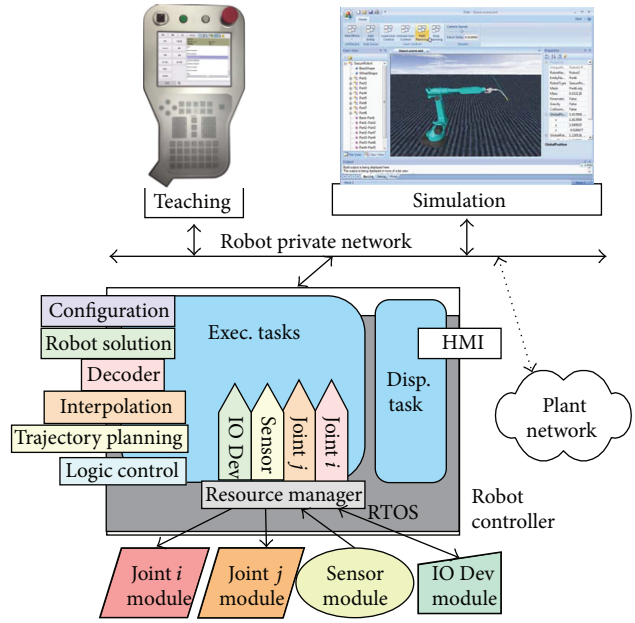


Figure 15: System structure of MWR.

coupling. The reuse cannot be implemented until the package is broken.

The robot joints are separated from the robot structure logically, that is, setting the isolated information to every joint. Then by reading the bus servo driver module, the single joint information is asserted.

The robot configuration formed by the joint connection can be expressed as D-H parameters matrix. Since the robot solution of 6-DOF joint robot is only influenced by D-H parameters, the D-H parameters are picked up as configuration file to support the configuration component.

(c) Implementation of MWR System. 6-DOF joint robot is a common industrial robot, and the overall system design has been more mature. However, its hardware and software do not use a unified system, and thus the compatibility with each other is quite bad. Current 6-DOF joint robot is generally designed and developed separately from scratch. The 6-DOF joint robots are composed of various of system structures, such as PC controller + motion control card, and an embedded controller + DSP, but for the open reusable software considering, the 6-DOF joint robot system structure should be single controller with pure software.

The single controller with pure software means that with the enough CPU capability, a single controller can process motion control, kinematics solver, and path interpolation functions entirely by software, and real-time OS running on the CPU will guarantee the real-time control of the underlying system. With this structure, system software and hardware can be reused in the best way; therefore here we design 6-DOF joint robot in this pure software structure.

The modular WMR system in this paper mainly includes 6 joint bodies, controller, joint drivers, decentralized control card, and system source and appendix. To be decoupled, the

controller outputs the real-time motor control messages such as position and velocity, and so forth; then the messages are sent to the joints via the decentralized control card, and motion control card such as P-MAC is not needed additionally.

In software the design requirement of 6-DOF joint robot in this paper is to track the robot end's trajectory. The system applies the real-time OS, which can ensure the hard real-time capability of the task operations, and the non-real-time modules, such as teaching, computing, and simulating, are distributed in network. As Figure 16, the user inputs the working task to the MWR system and interacts with the *Teaching Module* and *Simulation Module*, and then the working procedure is settled. The code handle node loads the decoding component, interpolation component, logical control component, data process component, and so forth. The logical task node executes the logical task, and meanwhile the axis control node executes the task of robot end's trajectory. To be clearly expressed, some data structures are simplified in the figure.

The thick lines with arrows express the hard real-time dataflow. With the scheduling of real-time system, the servo cycle time of axis position is settled 4 ms, and this can meet the requirement of automation workshop. In the design of this paper, *Teaching Module*, *Simulation Module* of MWR have been decoupled by network and the code handle node, logical task node, and axis control node are decoupled in software. The MWR hardware modules are reusable; in this design the axis can be replaced by different manufacture.

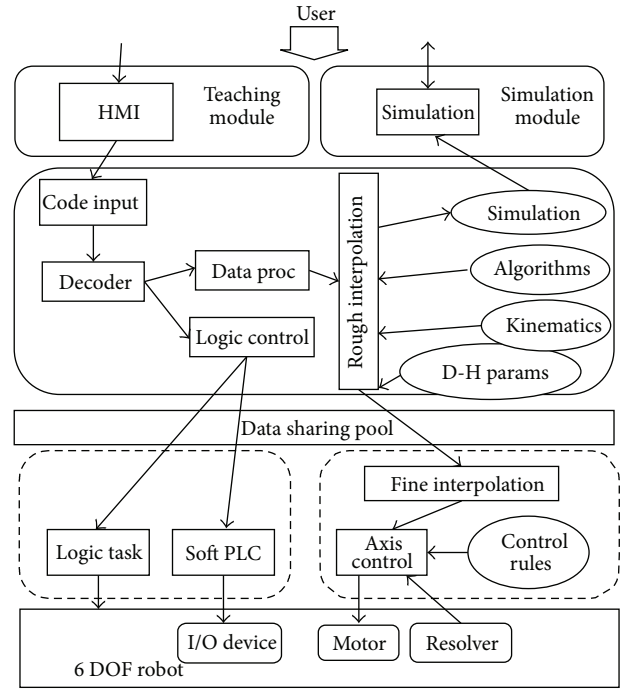


Figure 16: Dataflow chart of MWR.

7.2. Service Application—Medical Accompany Robot

(a) Introduction of Medical Accompany Robot. Medical Accompany Robot (MAR) is a differential wheeled service robot (Figure 17), and it is used for patients medical accompany, medical information interaction, and so forth.

MAR's hardware modules include Camera Module, Medical Sensor Module, Laser Senor Module, Infrared Sensor Module, Sonar Sensor Module, Robot Controller Module, Motor Driver Module, Monitor Module, and Operating Module.

Applied software modules include Video Processing Module, Control Algorithm Module, Position Perception Module, Human-Machine-Interaction Module, and Motor Control Module.

Applied system resource situations are the following:

Processor: Intel Atom Processor Z510 1.6 GHz

Memory: DDR2 1.0 GB

Data Storage: 8 GB CF Card

Operating System: Hybrid system RobOS [28] with RGMP

(b) Modular System Implementation of MAR. In F-Model framework, the MAR system is designed. Firstly MAR system is decoupled into hardware and software modules described in Section (a) and then register the *Camera Module*, *Medical Sensor Module*, *Laser Sensor Module*, *Infrared Sensor Module*, *Sonar Sensor Module* as *Sensor Class*, register the *Motor Driver Module* as *Driver Class*, register the *Robot Controller Module* as *Controller Class*, and register the *Monitor Module* and *Operating Module* as *Other Class*. The layer of hardware is simple, and the centralized structure is applied; that is, all sensor modules and motor driver modules are directly connected to the controller module, and all the medical accompany functions are implemented by software procedures running in controller module.

The sensor data process, *Main Control*, and *Motor Control* are settled as real-time nodes, while the *Video Process* and HMI are settled as non-real-time nodes. They run in the different areas of RobOS (Figure 18): the *Motor Control Node* and the *Position Percept Node* control the real-time device motor and sensors and run in the real-time OS area, while the *Video Process Node* and the *HMI Node*, respectively, connect to the *Camera Module* and *Monitor Module* and run in Linux area. The parts illustrated by thick lines form the real-time control chains of F-Model. To be clearly expressed, some data structures are simplified in the figure.

The communication between nodes is specified, and if the requirements are changed, the device modules need not to be changed, and the developer is to choose or design the necessary control algorithm node based on the requirement. The robot IDE supplies functions of choosing the MS software modules, developing control node procedures, setting real-time OS services, and so forth.

It can be seen that using F-Model, the MAR applications, system components, and hardware modules are decoupled, resulting in that the reuse is implemented. Based on the

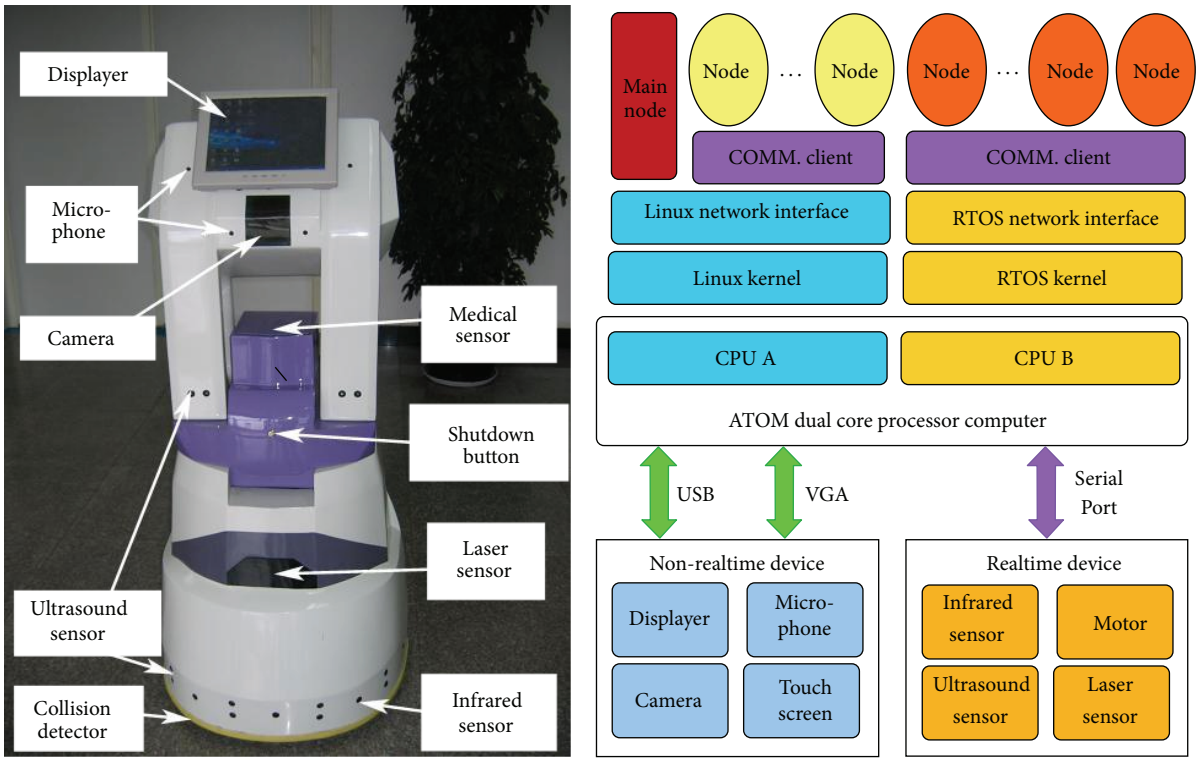


Figure 17: MAR system.

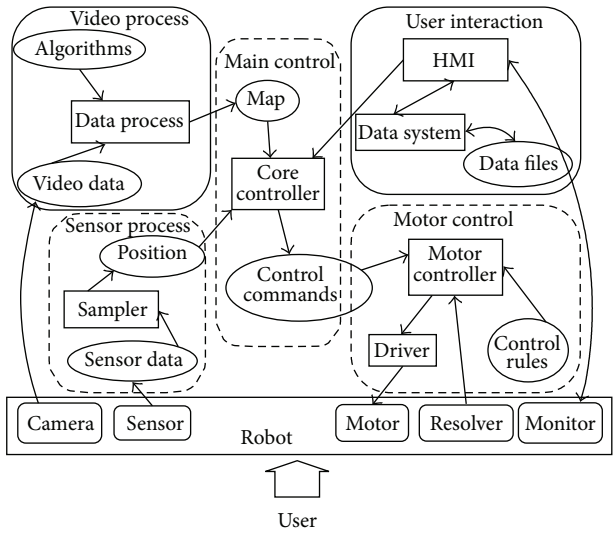


Figure 18: Dataflow chart of MAR.

requirements of MS control, the real-time working chains are reasonably arranged. MAR system runs well.

8. Discussion

Although a preliminary modular framework has already been constructed, there are some points to be discussed for the modular MS.

Firstly, the neat application node design depends on plenty of implementations of component or modules. If the various designs can follow the whole or the part of the same framework, these components will increase greatly. Future researchers should guarantee that the package of the valuable research results should be provided to others for reuse.

Secondly, uPnP is ideal future design requirement. MS requirement is various, thus the reasonable uPnP mechanism based on the specified interfaces needs further research. In this problem, the hardware module classification is primarily presented and the meaningful exploration is done for the further module specification.

Thirdly, the thoroughly automotive design is not implemented. Automotive design tool is the development aim of modular system; however, common automotive design tools' comprehensions for modular MS are inconsistent. These tools are not mature in automotive design and are hard to form complete modular design system.

Finally, F-Model does not limit the relationship between the MS itself and the surrounding devices and does not consider the module adjustment for that the system resource is extremely limited. It is just a preliminary exploration and further research and development are needed.

9. Conclusion

In this paper the core guidance for modular MS reuse and based on this a general framework is presented, which

includes software structure, hardware structure, and infrastructure; then two different kinds of modular MS instances are designed to establish the effectiveness of the framework. The MS modular design should follow the MS control information model, decouple the software and hardware, and assemble the MS applications with software engineering technologies.

Acknowledgments

This work is supported by Natural Hi-tech Development Program ("863") of China under the Research Project 2012AA041400. The authors greatly appreciate the researchers and engineers staffs of the ITR Laboratory in Robotics Institute of Beihang University for their tremendous help and kind assistance for them. This work was supported in part by the 863 Program of China under Grant 2012AA041400.

References

[1] N. Ando, T. Suehiro, K. Kitagaki, T. Kotoku, and W.-K. Yoon, "RT-Component object model in RT-Middleware—distributed component middleware for RT (Robot Technology)," in *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA '05)*, pp. 457–462, IEEE, Piscataway, NJ, USA, June 2005.

[2] N. Ando, T. Suehiro, K. Kitagaki, T. Kotoku, and W.-K. Yoon, "RT-Middleware: distributed component middleware for RT (Robot Technology)," in *Proceedings of the IEEE IRS/RSJ International Conference on Intelligent Robots and Systems (IROS '05)*, pp. 3555–3560, IEEE, Piscataway, NJ, USA, August 2005.

[3] H. Utz, S. Sablatnög, S. Enderle, and G. Kraetzschmar, "Miro—middleware for mobile robot applications," *IEEE Transactions on Robotics and Automation*, vol. 18, no. 4, pp. 493–497, 2002.

[4] D.-H. Choi, S.-H. Kim, K.-K. Lee, B.-H. Beak, and H.-S. Park, "Middleware architecture for module-based robot," in *Proceedings of the SICE-ICASE International Joint Conference*, pp. 4202–4205, IEEE, Piscataway, NJ, USA, October 2006.

[5] M. Mizukawa, S. Sakakibara, and N. Otera, "Implementation and applications of open data network interface ORiN," in *Proceedings of the SICE Annual Conference*, pp. 1340–1343, IEEE, Piscataway, NJ, USA, August 2004.

[6] H. Bruyninckx, "Open robot control software: the OROCOS project," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2523–2528, May 2001.

[7] N. Ando, T. Suehiro, K. Kitagaki, T. Kotoku, and W.-K. Yoon, "Composite component framework for RT-Middleware (Robot technology middleware)," in *Proceedings of the 2005 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM '05)*, pp. 1330–1335, Monterey, Calif, USA, July 2005.

[8] J. Jackson, "Microsoft robotics studio: a technical introduction," *IEEE Robotics and Automation Magazine*, vol. 14, no. 4, pp. 82–87, 2007.

[9] D. Brugali and P. Scandurra, "Component-based robotic engineering (part I)," *IEEE Robotics and Automation Magazine*, vol. 16, no. 4, pp. 84–96, 2009.

[10] D. Brugali and A. Shakhimardanov, "Component-based robotic engineering (Part II): systems and models," *IEEE Robotics and Automation Magazine*, vol. 17, no. 1, pp. 100–112, 2010.

[11] SONY. SONY Open-R Specification, http://www.aiboworld.tv/_download/common/OpenR.pdf.

[12] S. C. Ahn, J. H. Kim, K. Lim, H. Ko, Y.-M. Kwon, and H.-G. Kim, "UPnP approach for robot middleware," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1959–1963, Barcelona, Spain, April 2005.

[13] S. Magnenat, V. Longchamp, and F. Mondada, "ASEBA: an event-based middleware for distributed robot control," in *Proceedings of the IROS Workshop of the IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS '07)*, San Diego, Calif, USA, 2007.

[14] S. Magnenat, P. Rétornaz, M. Bonani, V. Longchamp, and F. Mondada, "ASEBA: a modular architecture for event-based control of complex robots," *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 2, pp. 321–329, 2011.

[15] G. S. Virk, "CLAWAR: modular robots for the future," in *Proceedings of the 3rd International Workshop on Robot Motion and Control (RoMoCo '02)*, pp. 73–76, 2002.

[16] Z. Ying, W. Tianmiao, W. Hongxing, L. Chengcheng, and L. Shiyi, "F-Model: model of modular robot based on functional component," in *Proceedings of the 15th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA '09)*, pp. 91–96, August 2009.

[17] H. Wei, S. Li, Y. Zou, L. Yang, and T. Wang, "A middleware based control architecture for modular robot systems," in *Proceedings IEEE/ASME International Conference on Mechatronics and Embedded Systems and Applications (MESA '08)*, pp. 327–332, December 2008.

[18] W. Hongxing, D. Xinming, L. Shiyi, T. Guofeng, and W. Tianmiao, "A component based design framework for robot software architecture," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '09)*, pp. 3429–3434, St. Louis, Mo, USA, October 2009.

[19] M. Törngren, D. Chen, and I. Crnkovic, "Component-based vs. model-based development: a comparison in the context of vehicular embedded systems," in *Proceedings of the 31st EUROMICRO Conference on Software Engineering and Advanced Applications (EUROMICRO-SEAA '05)*, pp. 432–440, September 2005.

[20] T. J. Van Beek, M. S. Erden, and T. Tomiyama, "Modular design of mechatronic systems with function modeling," *Mechatronics*, vol. 20, no. 8, pp. 850–863, 2010.

[21] I.-M. Chen and G. Yang, "Inverse kinematics for modular reconfigurable robots," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1647–1652, May 1998.

[22] J. Lei, T. Wang, and Y. He, "The modular approach based on functional components division for modular reconfigurable walking robot," in *Proceediongs of the ASME/IFToMM International Conference on Reconfigurable Mechanisms and Robots (ReMAR '09)*, pp. 540–544, June 2009.

[23] M. Khalgui, O. Mosbahi, Z. Li, and H.-M. Hanisch, "Reconfiguration of distributed embedded-control systems," *IEEE Transactions on Mechatronics*, vol. 16, no. 4, pp. 684–694, 2011.

[24] C. Zongyan and Y. Xinmin, "Reconfigurable controller for reconfigurable manufacturing system," *Computer Engineering and Applications*, vol. 39, no. 5, pp. 148–150, 2003.

[25] P. Mantegazza, E. L. Dozio, and S. Papacharalambous, "Rtai: real time application interface," *Linux Journal*, vol. 2000, no. 72, p. 10, 2000.

[26] RGMP project for RTOS and GPOS on Multi-Processor, http://rgmp.sourceforge.net/.

[27] Y. Tsuchiya, M. Mizukawa, T. Suehiro, N. Ando, H. Nakamoto, and A. Ikezoe, "Development of Light-weight RT-component (LwRTC) on embedded processor-application to crawler control subsystem in the physical agent system," in *Proceeding of the SICE-ICASE International Joint Conference*, pp. 2618–2622, kor, October 2006.

[28] M. Liu, Z. Shao, M. Wang, H. Wei, and T. Wang, "Implementing hybrid operating systems with two-level hardware interrupts," in *Proceedings of the 28th IEEE International Real-Time Systems Symposium (RTSS '07)*, pp. 244–253, Tucson, Arizona, USA, December 2007.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 913638, 8 pages
http://dx.doi.org/10.1155/2013/913638

Research Article

IPv6-Based Smart Metering Network for Monitoring Building Electricity

Dong Xu and Jingmeng Liu

School of Automation Science and Electrical Engineering, Beihang University, Xueyuan Road 37, Haidian District, Beijing 100191, China

Correspondence should be addressed to Jingmeng Liu; ljm@buaa.edu.cn

Received 8 April 2013; Revised 3 June 2013; Accepted 4 June 2013

Academic Editor: Yong Tao

A smart electricity monitoring system of building is presented using ZigBee and internet to establish the network. This system consists of three hardware layers: the host PC, the router, and the sensor nodes. A hierarchical ant colony algorithm is developed for data transmission among the wireless sensor nodes. The wireless communication protocol is also designed based on IPv6 protocol on IEEE 802.15.4 wireless network. All-IP approach and peer-to-peer mode are integrated to optimize the network building. Each node measures the power, current, and voltage and transmits them to the host PC through the router. The host software is designed for building test characteristics, having a tree hierarchy and a friendly interface for the user. The reliability and accuracy of this monitoring system are verified in the experiment and application.

1. Introduction

Environmental problems caused by human energy consumption are the major international issue faced by all countries worldwide [1]. The struggle between global warming and human beings is well recognized by the international society. Electricity is the main form of energy. As it is generated from fossil fuels, electricity represents a significant portion of carbon emission [2]. From the view of the distribution of power consumption recently, the proportion of urban residents and commercial electricity are very important. So how to save this part of energy consumption is an important issue to protect the environment [3, 4].

Monitoring the electricity condition of residents, especially the power consumption situation of some big buildings, is an effective method for saving energy. This approach was limited by technology level before. In last decade, it has been growing rapidly for several reasons, such as the improvement of microelectronics technology, development in wireless sensor, and the networking protocol. A wireless sensor network (WSN) consists of a large number of lightweight sensor nodes having limited battery life, computational capabilities, storage, and bandwidth [5]. WSN is carried out, thanks to the development of embedded system, system on chip, wireless communications, and low-power technology [6]. WSN has brought a revolution in the field of information perception for its low power, low cost, and the advantages of distribution and self-organization [7]. The wireless sensor network applications, which are low data rate detection and control systems, do not ask for real-time data transmission and do not require high bandwidth, but often require very low power consumption [8].

To meet the requirements of WSN, ZigBee is a high awareness of the wireless sensor networks in some short-range wireless communications technology [9, 10]. Since it has advantages including low power, low cost, automatic dynamic network, and high security, ZigBee is very suitable for wireless sensor networks. ZigBee protocol meets the OSI seven-layer protocol model, that from the bottom to the top is composed of the physical layer, data link layer, network layer, application support sublayer, and application layer, but also has a security module to provide network security features [11]. The main advantages of ZigBee are its low cost and low energy consumption characteristics. These advantages make ZigBee ideal for monitoring, data collection, and analyzing in various smart grid applications [12].

The sensor nodes of metering and monitoring system sense a change in the environmental or physical quantity and transmit this data to the base station. Hierarchy-based

multipath routing protocols are used for nodes that build a hierarchical relationship in order to discover efficient multiple paths [5, 13]. Ant colony algorithm is mostly used to find the optimal path in a large number of nodes [14]. The base station is usually a powerful machine like a PC. Internet Protocol version 6 (IPv6) is needed for the PC in the network [15]. IPv6 is a version of the internet protocol (IP) that is designed to succeed Internet Protocol version 4 (IPv4) and is also the basic technology of internet [16]. IPv6 simplifies aspects of address assignment and network renumbering when changing internet connection providers [17]. Network security is also integrated into the design of the IPv6 architecture, and the IPv6 specification mandates support for a fundamental interoperability requirement [18]. However, the WSN architecture is different from the IPv6 network architecture. Achieving all-IP communication between WSN and IPv6 networks needs further researches [19, 20].

This paper presents a smart metering network system based on the IPv6 network protocol and ZigBee protocol for residential power measurement. Because wireless sensor nodes are located in the AC power outlets distributing throughout the whole building, the routers are used for data changing between host PC and the sensor nodes. The hierarchical ant colony algorithm is designed for data transmit ion among the wireless sensor nodes.

This paper is organized in six sections. The system structure is presented in the second section. Then, the wireless sensor node is developed in the third section. In Section 4, the software on host PC is designed. Section 5 illustrates the IPv6 network and the hierarchical ant colony algorithm based on ZigBee. Simulation and experiment results that show the validity of the presented are demonstrated approaches in Section 6.

2. System Structure

Figure 1 shows the system structure of the smart metering network. This monitoring system is composed of three hardware layers, which, from top to bottom, are the host PC, the router, and the wireless sensor nodes. Each layer has its software. The network protocol based on hierarchical ant colony algorithm is presented to improve the efficiency of data transmission among the wireless sensors. Between the host PC and the sensors, IPv6 protocol is achieved through the router.

The host PC is used for data analysis and power control. The software in the PC is implemented on the QT environment based on Linux operation system. First, the host software intelligently networks all the sensor nodes in the building. Then, according to the requirements of the user, it completes the configuration of each node. Each sensor node, based on the received configuration information, begins to upload data, to real-time protect the power, and to regularly maintain the system network.

The routers are set up between the host software and sensor nodes for data conversion between different protocols. Using a router in this structure also improves the stability of the network and transmission distance.

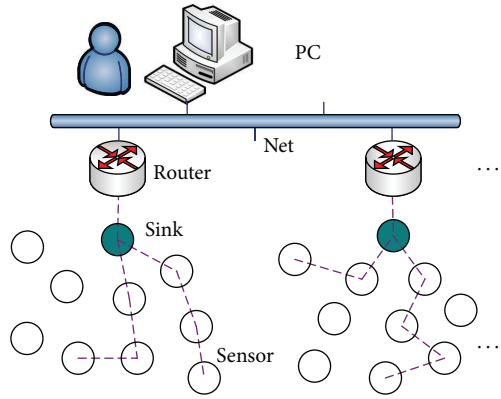


Figure 1: The network structure of system.

In the wireless sensor, the embedded software is developed based on an embedded operating system, Contiki is a multitasking open source operating system with a support of TCP/IP networks including IPv6. Contiki needs only a small amount of memory that only 2 kb of RAM and 40 kb of ROM are needed to provide multi-tasking environment and built-in TCP/IP support. It can dynamically load the upper application at runtime and achieve interprocess communication through the use of information events of lightweight process model. Because of these advantages, this operating system is very suitable to develop applications of sensor nodes.

The composition details of the smart metering network will be presented in the following.

3. Wireless Sensor Nodes

The sensor node consists of voltage and current sensors, ZigBee communication, power conversion, and relay control. The hardware structure of the node is shown in Figure 2. The node is set up by receiving the information from the host software that includes the sampling frequency, filter parameters, and power control parameters. After that, it measures the RMS of the voltage and current, the instantaneous values of voltage and current, active power, reactive power, and apparent power. Then, these data are transmitted to the host software according to the communication protocol.

All the functions of the sensor node are implemented in the microcontroller called STM32W. It has advantages such as supporting for wireless transmission protocol, convenient interface with the measuring sensor, powerful computing capabilities for data processing, on-chip program and data memory to save the board space, and IO port for power control. It has a 32-bit Cortex-M3 microprocessor, Flash and RAM memory, and peripherals to design 802.15.4-based systems [21].

3.1. Electricity Measurement. One main function of the sensor is to measure the voltage, current, and power. As the microcontroller does not have a measure model, an energy metering IC named ADE7753 is chosen for this purpose. In addition to RMS calculation and active and apparent

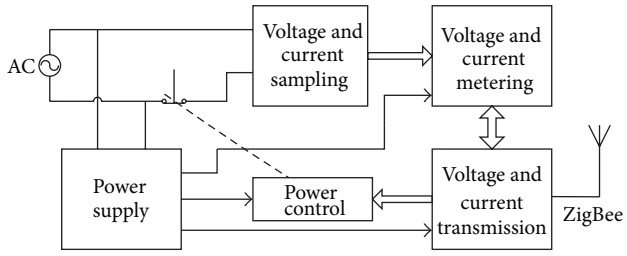


Figure 2: The block diagram of wireless sensor node.

Table 1: Data interface command.

Address	Command	Data	Description
0x01	WAVEFORM	24 bits	The sampled waveform data
0x02	AENERGY	24 bits	The active power data
0x05	VAENERGY	24 bits	The apparent power data
0x16	IRMS	24 bits	Current channel RMS value
0x17	VRMS	24 bits	Voltage channel RMS value
0x22	IPEAK	24 bits	The maximum current value
0x24	VPEAK	24 bits	The maximum voltage value

power information, the ADE7753 also accumulates the signed reactive energy. It has two fully differential voltage input channels, each of which has a PGA (programmable-gain amplifier) with possible gain selections of 1, 2, 4, 8, and 16. It also contains an on-chip power supply monitor continuously monitoring the analog supply [22]. This is useful to ensure correct operation during power up and down.

The voltage and current sensing circuits of the sensor node are designed according to the measurement signal interface characteristics of ADE7753. For voltage signal acquisition, a 1 MΩ resistor and a 1.2 kΩ resistor are connected in series, so that the voltage dropped on this 1.2 kΩ resistor is the sampling signal. Considering the sampling accuracy and system size and cost, current sampling uses the method of shunt resistor. A 5 mΩ resistor is set in the live ware to transform the output current into voltage drop to obtain the differential voltage signal for sampling.

3.2. Data Interface. The sampling data is read from the ADE7753 via the serial peripheral interface bus (SPI). In this full duplex data transmission, the microcontroller is the master, while ADE7753 is the slave. The serial clock for a data transfer is applied at the SCLK by the microcontroller. All data transfer operations are synchronized to the serial clock. Data is shifted into the ADE7753 at the DIN logic input on the falling edge of SCLK and shifted out at the DOUT logic output on a rising edge of SCLK. The SPI transfer mode based on interface timing with half-cycle delay can guarantee the stability of data transmission. The microcontroller reads data, including RMS, waveform, the power, and the peak value, from ADE7753 with the commands in Table 1.

3.3. Power Supply. Sensor nodes are used to measure the state of residential electricity consumption whose power supply voltage is AC 220 V. The circuit of sensor node includes a microcontroller STM32W and an energy measurement chip ADE7753, so these circuits require DC 5 V and 3.3 V power supplies. A reasonable approach is to get these DC powers from AC 220 V. There are two technical solutions, switching power and linear power, that can transform AC to DC power. The circuit of switching power is complex, has high cost, and has a strong electromagnetic interference to the measurement circuit. The voltage fluctuation of residential electricity is less than 10%. So a linear power supply can meet the system requirements of DC power to affect the stability of the operation of the microcontroller and, more importantly, to affect the measurement accuracy of the voltage, current, power, and other energy information. In the design process, when using an external DC power supply, the data is from 2 to 3 bits more accurate than that using the 220 V linear power circuit. The DC power filter, particularly the π-filter, is an effective way to solve these problems.

3.4. Alarm and Control. The function of alarm and power on and off control is needed in this smart metering system. Because different appliances have different power requirements, the alarm of system is not the same measurement limit. Each wireless sensor node has received the settings from the host software to set the protection limit of the power of the node. The connected electrical equipment using power plug generally does not change, so this limit value preset method is used for the first time when the equipments are connected. When the node power is over the setting limit for a certain time such as 5 seconds, an alarm message is sent out. According to the value of the power, the node decides to take the initiative to cut off power supply immediately or after waiting for the power control commands. For the case of electric short circuit caused error, the node will judge the error according to the measured current value. If there is an error, it will cut off the power to prevent accidents and alarm this information at the same time.

A switch is required to control the power on and off. Taking into account the state of normal power supply, only the abnormal condition or operation of the command issued by the host software will cut off the power supply. For monitoring and controlling power supply of the residents, switching operation is not frequent; the switch relay is selected to meet the system requirements. In normal operation, the relay coil controlled by the transistor is off, and the normally closed contact is used as the power control switch. When relay coil is energized, the contact is off to cut off the power.

3.5. Embedded Software. The embedded software of sensor node is developed fully using the IPv6 network protocol support of the operating system. According to the characteristics of the metering network, the function of this embedded software is as follows: first, each node with automatic search and networking capabilities has the ability to dynamically create a metering network; second, as the main task is to measure the power, the node should ensure data accuracy though there may be outside interference in the case of processing; according to the protocol, there are also the tasks to achieve the state control and power switch control under the relevant command of the host system.

A simple description is given of the functions and work processes of this embedded software as follows. The node initialization function sensorInit( ), including the ADE7753 initialization function embInit7753( ), is the beginning of the software work. When the node is ready, the network initialization function embNetworkInit( ) is to set up the metering network. After the networking process, according to the feedback of network state function embNetworkState( ), two functions, embSendData( ) and embReadData( ), begin to receive data, send data, and analyze protocol. The function applicationTick( ) contains the node applications, such as parsing command and control. For example, the functions readOperation( ) and writeOperation( ) are used to control ADE7753 and to read the measurement data.

4. Software on Host PC

All the data are transmitted to the PC for user management. The software in host PC should have the following functions: (1) data collection and setting; (2) energy analysis, including energy consumption statistics, limited power consumption control, and energy distribution; (3) monitoring device control and protection. The software is named as iSEnergy. The software logic is designed in different levels for hierarchical management. It should consider different rooms in the building and the outlets in the room. The main layers are as follows: (1) for a building, it is able to view the energy usage of the entire building and each room, and, it can analyze and control the energy usage of each room; (2) for a room, it is able to view the energy consumption of the room and each equipment, and it could also control each device in the room; (3) for a monitoring device, it is able to view and control the energy usage.

The host software and wireless sensor nodes communicate using IPv6 wireless network. Relative to the ZigBee protocol, the use of IPv6 is simpler for system control and can achieve point-to-point access between the nodes.

Taking into account the cross-platform software features, iSEnergy is written using Qt, an open source software library. Qt is a well-known GUI Linux platform graphics library, with rich interface functions and stable system architecture. iSEnergy is divided into several functional modules as follows:

(a) network communicating module: monitoring the IPv6 network communications between devices; packing and unpacking work based on user-defined data protocol;

(b) database communicating module: storing the data of the database; monitoring data saved for the statistics and analysis;

(c) user interface module: core of the system module responsible for interacting with users and other data to communicate between modules.

The main program starting process is shown in Figure 3. The running start interface, the main interface, the configuration interface, and the monitoring interface are shown in Figure 4.

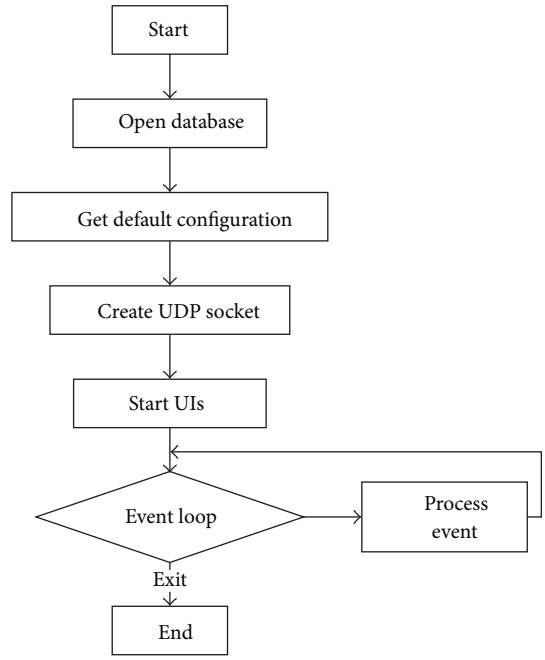


Figure 3: The program starting process of host software.

5. Network of System

5.1. IPv6 Network Protocol. ZigBee is the protocol of wireless sensors. IEEE802.15.4 is responsible for the physical layer and media access control layer standard of this protocol. The IPv6 network cannot be directly built on IEEE802.15.4 network; this problem should be solved in the network communication protocol. There are two methods, all-IP approach and peer-to-peer mode, that can achieve ZigBee wireless sensor nodes access in an IPv6 network.

The all-IP approach is an address-centric way for data-centric sensor nodes to solve communication problems, so it will reduce efficiency and increase the power consumption of communication nodes. If individually visiting each node in the case, these nodes should have a global unique IP address. The all-IP approach provides strong support for the entire communication protocol.

Peer-to-peer mode achieves the interconnection between the inside and outside networks by setting a particular gateway to convert the protocol between WSN and IPv6 protocol in the same layer. In accordance with the work in different layers, the gateway can be divided into application-level gateway and network address translation gateway. The drawbacks are low user transparency, difficulty in using a variety of services offered by WSN, and the difference between the network protocols.

The smart metering network is designed by integrating these two modes. Each sensor node supports the IPv6 protocol, but also designed a gateway to solve transmission distance limitations of ZigBee. The gateway node as the interface of internal network and the user terminal can temporarily store and forward the data. Because the sensor node supports IPv6

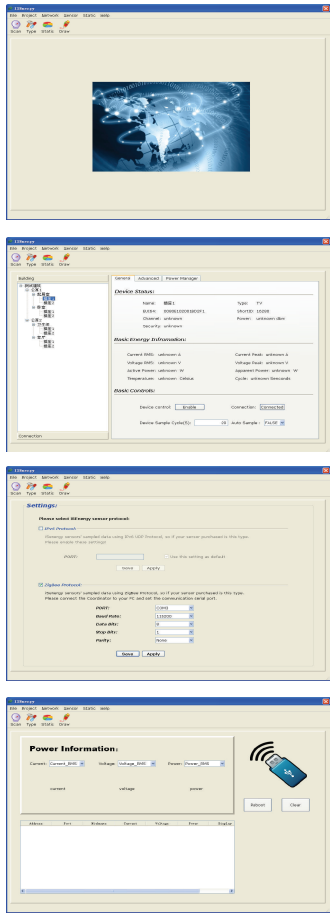


Figure 4: The interfaces of iSEnergy.

protocol, the network with each node having its own address has good transparency.

5.2. Wireless Protocol. An intelligent algorithm to find the optimal path of the wireless sensor network is an important direction for wireless sensor networks. A hierarchical ant colony algorithm is presented as the communication protocol of the wireless sensor network. As shown in Figure 5, the 1st and 2nd nodes are in the first layer, both of which are on transmission range of the 3rd node. So, if data is translated from the 3rd node to the sink node, there are two most efficient paths off-line optimization goal is to find out the lowest energy consumption path. Therefore, the distance which could substitute the energy consumption between two nodes in the simulation is calculated. Through the distance of each path, the optimum next node can be found out. If it is busy, the other node could be selected.

Ant colony algorithm is used to find the optimal path in a large number of nodes. Because timely reaction to the electricity state of the household appliances is important, the efficiency is put in the first place instead of the energy consumption. Firstly, the clustering method is used to identify the efficient paths. And then the ant colony algorithm is used to find the optimal path in these paths.

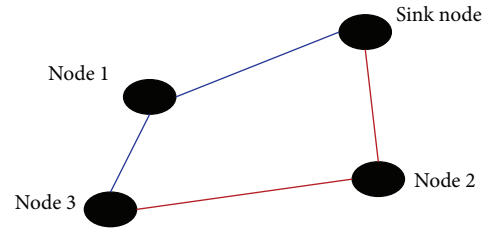


Figure 5: The nodes in the wireless network.

The algorithm is approached as follows.

Step 1. Report the "dead center" which cannot interact with other nodes so that the users can modify their places or other activities.

Step 2. Let the sink node be level zero and initialize the ants. The ants take the sink node as current position and search through the range of the movement of the ants to find the first layer nodes. The next optimization nodes which exist in all first layer nodes are the position of sink are saved.

Step 3. Initialize the ants, and let them take every first node as current position. Then search through the range of the movement of the ants to find the second layer nodes. When finding them, (1) these nodes must exist in the range of the movement of the ants; (2) the nodes should not belong to the set of the nodes of the first layer. Finding the following nodes is the same as the second layer nodes: (1) the nodes must be in the range of the movement of the ants; (2) the nodes should not belong to higher layer nodes.

Step 4. By finishing every step, we could get the following purposes. Every node can save information including which layer does this node belong to and the next nodes' IP. Then we can choose the optimal path to interact with the rememorized next node. If the better node which we choose is busy in working or breaking down, we should choose another next node to interact with the next one. Then, this node becomes the current node to interact with the next one. Finally, it reaches to the sink node.

By finishing the above steps, the path has reached the highest transmission efficiency.

6. Experiment and Application

6.1. Simulation and Experiment. In the simulation, a theoretical model is built with the parameters. The location (x, y, z) is used to substitute IP. The transmission distance (d) is the distance between two nodes. For the transmission efficiency (V): between two nodes, although the distance is different, but relative to its transmission in terms of speed, time difference of distance is negligible, so the transmission efficiency is not affected by the two-node distance, but by the number of nodes. For the two-way interaction, the actual wireless sensor is bidirectional.

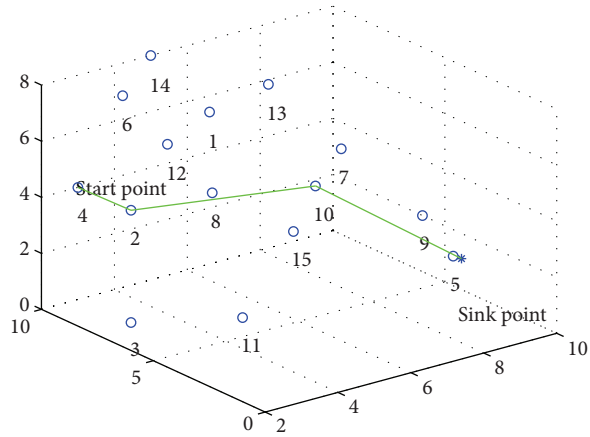


Figure 6: The path after stratified.

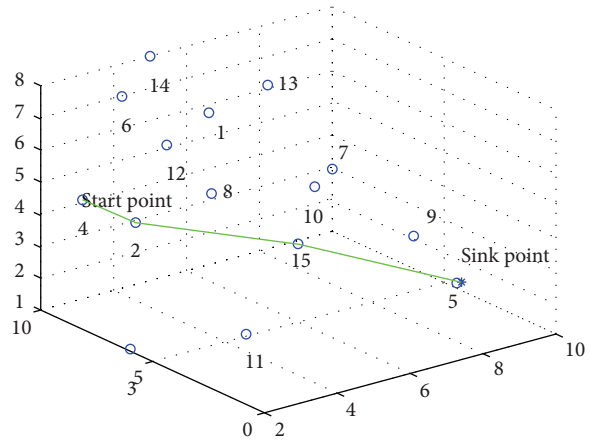


Figure 7: The optimized path of WSN.

The conditions of the optimal path are as follows: firstly, the transmission must be successful; secondly, the transmission efficiency should be excellent; finally, the transmission energy consumption should be low.

The high efficiency of transmission path after the stratification in the system of a total of 15 nodes is shown in Figure 6. The starting point is node 4. The data reach to the sink node passing through node 2 and node 10. In this path, node 4 belongs to the third layer, node 2 belongs to the second layer, and node 10 belongs to the first layer.

After optimization by the ant colony algorithm, the path is shown in Figure 7. The starting point is node 4, passing through node 2 and node 15, to reach the sink node, in which node 4 belongs to the third layer, node 2 belongs to the second layer, the node 15 belongs to the first layer. Compared with the first path, this path is not only efficient, but also has low power.

In the testing experiment, the smart metering network is installed in a building, a simple environment with less interference. Under the condition that the antenna is installed and that there are no obstructions, the maximum transmission distance is not more than 70 meters to ensure the normal communication between router node and end node. When obstructions exist, in order to ensure the reliability of wireless communications and to analogize the system installed in the building environment, it is more reliable in the case that the distance between nodes does not exceed 20 meters.

By the installation space limitations, if the antenna is not used, the reliable transmission distance between nodes is 10 meters. When the transmission distance is ensured, the ZigBee router node can achieve good control of the terminal nodes, and the terminal nodes can feed back information in time. Because the system communication delay is very short, usually for the dozens of subtle, it meets the measurement requirements very well.

The measurement accuracy, as well as the stability of the network of the metering system, is an important factor for application. Therefore, we tested and calibrated the accuracy and linearity of the voltage, current, and power measurement of sensor node. Take the current measurement, for example, the results of the experiment are given below. A sliding rheostat is connected to an autotransformer, so we get that a circuit can continuously adjust the supply voltage and operating current.

In the current measurement experiment, the currents were adjusted to 0 A, 2.5 A, 5 A, 7.5 A, and 10 A, and the data acquisition time interval is 1 s. When the current is 0 A, the test result is not zero, so the mean of the measured data is taken as the zero-off-set calibration. The current measurement results after this calibration are shown in Figure 8 with good accuracy. Figure 9 shows the discreetness of the current data when it is 5 A. The data range is between 4.98 A and 5.04 A, and the data fluctuation is small. Figure 10 shows the relationship of the means of collected current data. It can be seen that the various measurements have good linearity. The experiment shows that the deviation of the sensor node data collection ranges within 0.8 percent; after mean filtering, data accuracy is higher than 0.1 percent, so the accuracy can well meet the energy measurement precision.

6.2. Application Results. The smart metering network designed in this paper is exquisite and stores the electrical data of the electrical equipment, buildings, rooms, and the entire building. A different analysis model of the range of electricity consumption is set up to monitor the state of electricity. This metering network is installed in the Electrical and Electronic Center of Beihang University to monitor electrical conditions in a student laboratory.

The host software is installed on a computer connected to the Internet and a router is connected to this host computer through the Internet. The sensor node is installed in the installation box in the wall, which will not bring any impact to the experimental environment. The sensor node connects to the host computer through a router with Zigbee wireless network.

After all the sensors are powered to run, the electrical equipments are connected to the installation of the sensor node power box and would not change this relationship in this application. 30 sensor nodes were installed in an experiment laboratory, each of which monitored an electrical device in this application. The electrical equipment included

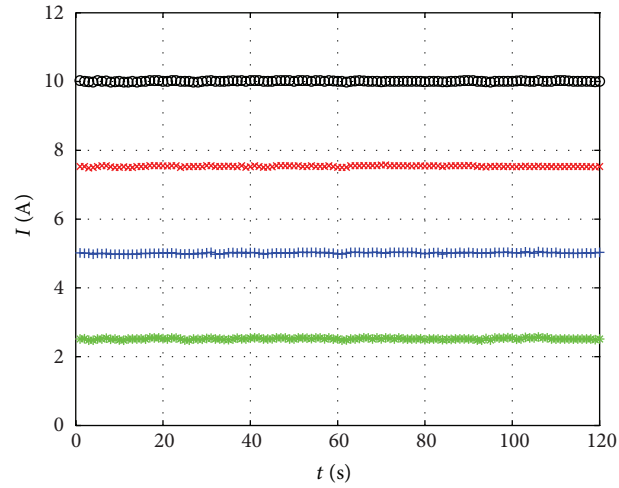


Figure 8: Current measurement data.

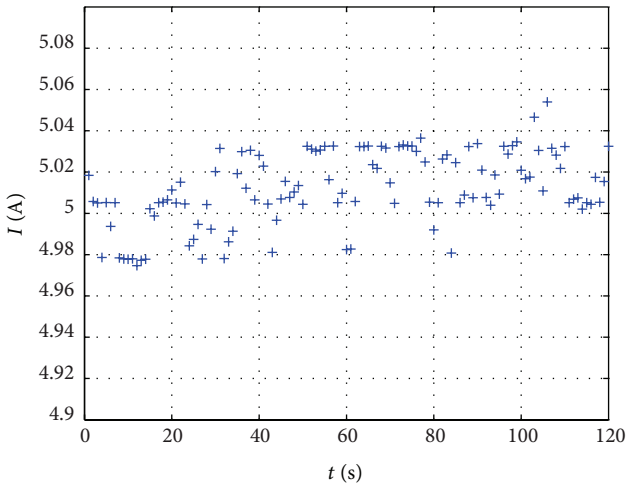


Figure 9: Fluctuation data.

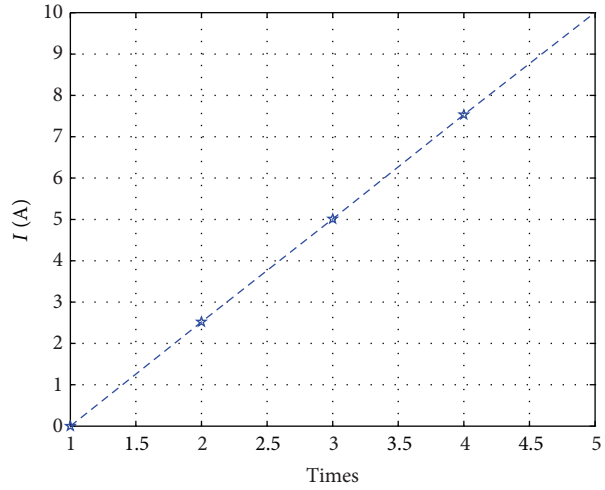


Figure 10: Linearity data.

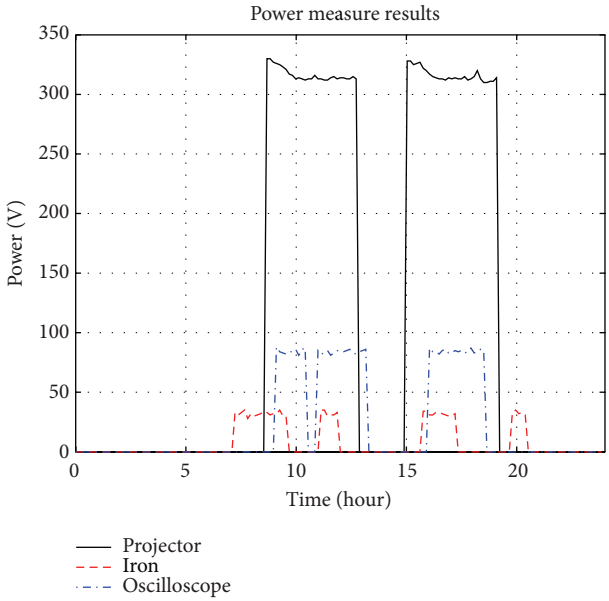


Figure 11: Power curves of some equipments.

computers, projectors, oscilloscopes, signal generators, soldering iron, electrical test platform, and so on. Sensor nodes worked 24 hours a day monitoring the power consumption. The host software was set to record the data of all the nodes in a week. During this week, the system runs stably, and 30 sensor nodes continuously worked to collect metering data and to communicate through the wireless network with the host.

Each sensor node is connected to a device. The sensors sample the data including voltage RMS, current RMS, and power consumption. On the host software, all of the devices are monitored and can be controlled at any time.

Figure 11 shows the power measurement curves of a projector, a soldering iron, and an oscilloscope. The projector, whose measured result is 310 W, consumed the most power in these electrical equipments. When using in the laboratory, the teachers who forget to turn off the projector after use will waste more energy. The power consumption of oscilloscope was about 80 W, while that of the iron was 30 W. The soldering iron was turned off when not used, and the scope worked a long time in the whole process.

7. Conclusion

A smart metering network system is designed in this paper which contains three layers of hardware: the wireless sensor nodes, the router, and the host PC. The system has two levels of protocols. One is the IPv6 network that is between the PC and the sensor nodes, and the other is the hierarchical ant colony algorithm among the sensor nodes based on Zigbee. This system achieves intelligent building monitoring of all electrical equipments.

With no additional installation requirements, the sensor nodes are installed in the building walls in the terminal

box. These sensor nodes, including ADE7753 as energy measurement core module, can take accurate real-time monitoring through circuit design and algorithm processing. The measurement accuracy of voltage, current, and power is more than one-thousandth. The microprocessor used in the sensor node integrates Zigbee communication module. So an IPv6 network, in which each node achieves a separate IP address, is established based on Zigbee protocol using the routing node to improve the transmission distance. The data transmission distance between sensor nodes is within 10 meters with no antenna installed, and this can meet the monitoring requirements for the building. The host software of smart monitoring network system is achieved in a PC, and it displays and analyzes power monitoring data with a friendly and intuitive interface acquisition. The smart system improves the safe use of electricity by the effective measures of active power control and automatic overload protection function. Experimental results verify the feasibility of the system, and the application in the office building of the Electric and Electronic Center of Beihang University obtains satisfactory results.

Conflict of Interests

The authors declared that they have no conflict of interests in this work.

Acknowledgment

This work is supported by the National Nature Science Foundation of China (under Grant 61203353).

References

[1] V. Cecchi, X. Yang, K. Miu, and C. O. Nwankpa, "Instrumentation and measurement of a power distribution system laboratory for meter placement and network reconfiguration studies," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 4, pp. 1224–1230, 2007.

[2] S. S. Chavan, S. Jayaprakash, and V. J. Kumar, "An open standard protocol for networking of energy meters," *IEEE Transactions on Power Delivery*, vol. 23, no. 4, pp. 1749–1753, 2008.

[3] C. Bennett and S. B. Wicker, "Decreased time delay and security enhancement recommendations for AMI smart meter networks," in *Proceedings of the Innovative Smart Grid Technologies Conference (ISGT '10)*, January 2010.

[4] J. C. S. de Souza, M. B. Do Coutto Filho, M. T. Schilling, and C. de Capdeville, "Optimal metering systems for monitoring power networks under multiple topological scenarios," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1700–1708, 2005.

[5] K. W. Sha, J. Gehlot, and R. Greve, "Multipath routing techniques in wireless sensor networks: a survey," *Wireless Personal Communications*, vol. 70, no. 2, pp. 807–829, 2013.

[6] L. Cao, W. Jiang, and Z. Zhang, "Automatic meter reading system based on wireless mesh networks and SOPC technology," in *Proceedings of the 2nd International Conference on Intelligent Networks and Intelligent Systems (ICINIS '09)*, pp. 142–145, November 2009.

[7] E. Valigi and E. Di marino, "Networks optimization with advanced meter infrastructure and smart meters," in *Proceedings of the 20th International Conference and Exhibition on Electricity Distribution (CIRED '09)*, June 2009.

[8] M. Bsoul, A. Al-Khasawneh, A. E. Abdallah et al., "An energy-efficient threshold-based clustering protocol for wireless sensor networks," *Wireless Personal Communications*, vol. 70, no. 1, pp. 99–112, 2013.

[9] T. Maity and P. S. Das, "Intelligent online measurement and management of energy meter data through advanced wireless network," in *Proceedings of the International Conference on Devices and Communications (ICDeCom '11)*, February 2011.

[10] H. Y. Tung, K. F. Tsang, and K. L. Lam, "ZigBee sensor network for advanced metering infrastructure," in *Proceedings of the International Conference on Consumer Electronics (ICCE '10)*, pp. 95–96, January 2010.

[11] N. C. F. Tse, W. H. Lau, and J. Y. C. Chan, "ZigBee based smart metering network for monitoring building integrated electric vehicle charging circuits," in *Proceedings of the IEEE PES General Meeting (PES '10)*, July 2010.

[12] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. Culler, "Design and implementation of a high-fidelity AC metering network," in *Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN '09)*, pp. 253–264, April 2009.

[13] G. Wood, "IPv6: making room for the world on the future Internet," *IEEE Internet Computing*, vol. 15, no. 4, pp. 88–89, 2011.

[14] J. Lee and G. L. Park, "Power load distribution for wireless sensor and actuator networks in smart grid buildings," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 372982, 8 pages, 2013.

[15] J.-C. Liu and C.-D. Chung, "Distributed estimation in a wireless sensor network using hybrid MAC," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 3424–3435, 2011.

[16] S. S. Kolahi and P. Li, "Evaluating IPv6 in peer-to-peer 802.11n wireless LANs," *IEEE Internet Computing*, vol. 15, no. 4, pp. 70–74, 2011.

[17] M. Mackay and C. Edwards, "A managed IPv6 transitioning architecture for large network deployments," *IEEE Internet Computing*, vol. 13, no. 4, pp. 42–51, 2009.

[18] J.-H. Lee, "Cross-layered IPv6 neighbor discovery scheme over WLAN mesh networks," *IEEE Communications Letters*, vol. 13, no. 12, pp. 992–994, 2009.

[19] C. Vogt and M. Zitterbart, "Efficient and scalable, end-to-end mobility support for reactive and proactive handoffs in IPv6," *IEEE Communications Magazine*, vol. 44, no. 6, pp. 74–82, 2006.

[20] M. Park, J.-T. Kim, E.-H. Paik, and K.-R. Park, "Deployment strategy and performance evaluation of the IPv6 home network using the home server," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 1, pp. 114–119, 2007.

[21] X. N. Wang and H. Y. Qian, "Research on all-IP communication between wireless sensor networks and IPv6 networks," *Computer Standards & Interfaces*, vol. 35, no. 4, pp. 403–414, 2013.

[22] S. Jianhua and H. Liping, *Principles and Applications of STM32W Microcontroller With ZigBee Radio Frequency*, Press of Beihang University, 2010.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 268046, 7 pages
http://dx.doi.org/10.1155/2013/268046

Research Article

Extended-Maxima Transform Watershed Segmentation Algorithm for Touching Corn Kernels

Yibo Qin,[1] Wei Wang,[2] Wei Liu,[1] and Ning Yuan[1]

[1] *School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255049, China*
[2] *College of Engineering, China Agricultural University, Beijing 100083, China*

Correspondence should be addressed to Wei Liu; weikey@sdut.edu.cn

Received 22 March 2013; Accepted 4 June 2013

Academic Editor: Yong Tao

Touching corn kernels are usually oversegmented by the traditional watershed algorithm. This paper proposes a modified watershed segmentation algorithm based on the extended-maxima transform. Firstly, a distance-transformed image is processed by the extended-maxima transform in the range of the optimized threshold value. Secondly, the binary image obtained by the preceding process is run through the watershed segmentation algorithm, and watershed ridge lines are superimposed on the original image, so that touching corn kernels are separated into segments. Fifty images which all contain 400 corn kernels were tested. Experimental results showed that the effect of segmentation is satisfactory by the improved algorithm, and the accuracy of segmentation is as high as 99.87%.

1. Introduction

1,000-kernel weight is an important indicator of seed vitality. This parameter used to be measured by hand, but the process is inefficient, and the result is inaccurate. Using digital image processing to automatically measure 1,000-kernel weight is a fast and effective method being developed for future use with corn seed [1]. As the kernels being imaged are usually touching each other, the counting accuracy by digital image processing is not as high as required, hence the need to develop an improved algorithm for the segmentation of the touching corn kernels.

Most of the literature about digital image segmentation of large quantities of seed focuses on round-shaped grains, such as rice and wheat [2–4]. For corn kernels, which have various sizes and irregular shapes, it is more difficult to segment the image, and there are few segmentation algorithms for corn kernels. Yaqiu et al. [1] proposed a watershed transform with gradually changing threshold for the segmentation of corn seed in digital images. That method is based on the area and the average width of the corn kernels, its main shortcoming is the large amount of calculation required. Xun et al. [5] presented an approach based on intersection areas and contours of touching kernels to search for segmentation points. This algorithm is most effective for two touching kernels. However, when the images show a large number of seeds with complicated connections, the authors report that the effect of segmentation is unsatisfactory.

Present cereal image segmentation methods include mathematical morphology [2], the concave point detection algorithm [6, 7], elliptic curve fitting [8, 9], the distance transform and watershed segmentation algorithm [10], the kernels' contour curvature [11], and the active contour algorithm [12]. As the watershed transform method has a good response to weak edges and smooth segmentation curves, it has widespread applications.

The watershed transform is a segmentation algorithm which is based on the topological structure of an image and can divide the image into a plurality of regions. The algorithm originated in the field of geodesy, specifically the study of watershed areas. The watershed area is digitally imaged and rendered into grayscale, with darker colors indicating lower altitudes. A water basin is thus constituted by a group of pixels surrounding a local grayscale minimum, including the minimum itself. With increasing water level, the region surrounding each local grayscale minimum gradually

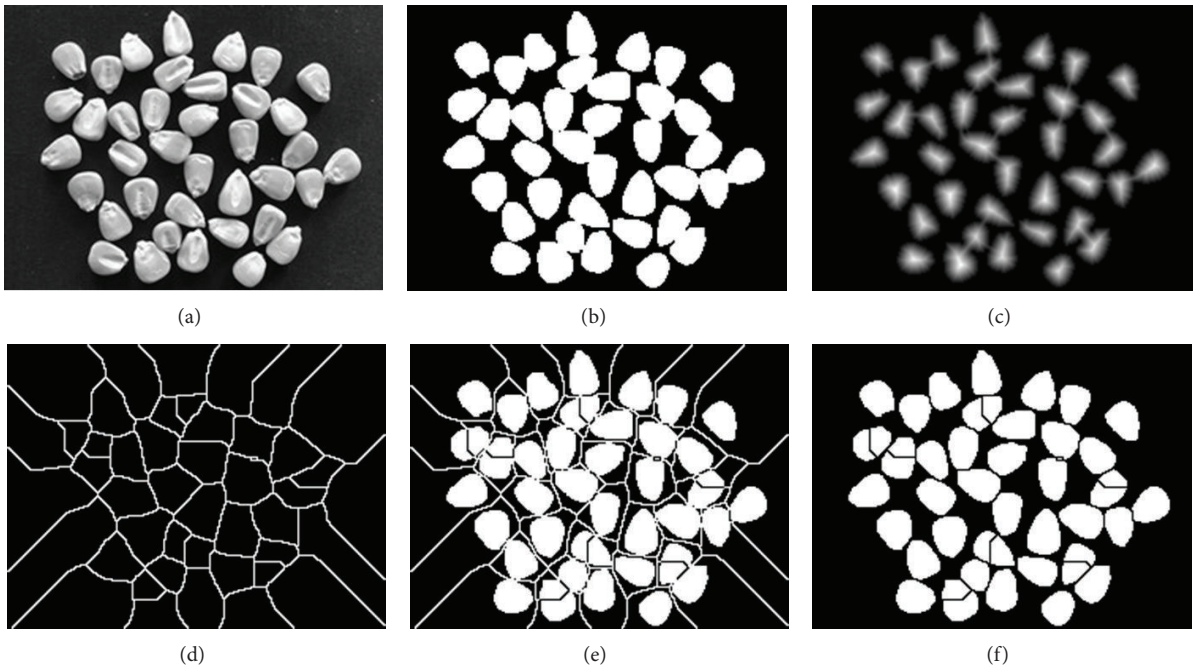


Figure 1: Watershed segmentation transform steps: (a) original grayscale image, (b) binary image, (c) distance-transformed image, (d) watershed ridge lines, (e) original binary image overlaid with watershed ridge lines, and (f) complementary image of watershed ridge lines superimposed on the binary image.

expands outwards, eventually forming a junction between neighboring water basins, called a watershed boundary [13, 14]. Corn kernels also have indentations due to their irregular shape, which function analogously to the water basins in the algorithm, allowing the "watershed boundaries" in the image, each surrounding a single kernel, to be computed.

Because of the irregular shape of corn kernels, the traditional watershed transform method always leads to oversegmentation. In this paper, by isolating the reason for the oversegmentation, we propose an improved watershed segmentation algorithm based on the extended-maxima transform that solves the problem of oversegmentation of touching corn kernels.

2. Segmentation Algorithm

2.1. Watershed Segmentation Transform. In this paper, all transformations are based on binary images. Before applying the watershed transform, we need to first carry out the distance transform. The distance transform (DT) labels each pixel of the image with the distance to the nearest boundary pixel in a binary image [15]. This transform correctly segregates round and simple touching objects, but for kernels with irregular shapes, the images usually show several maxima after the distance transform, which leads to oversegmentation.

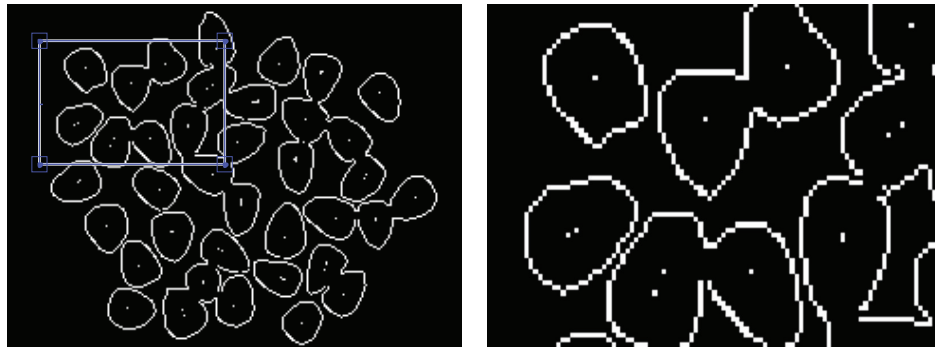
Image segmentation processing is shown in Figure 1, where image (a) represents the original image in grayscale. The image is converted to binary image (b) by morphologically opening [16] to smooth the edges. The image is subjected to a distance transform as shown in image (c). The grayscale image is equivalent to a topological map, with the maxima as peaks and the minima as valleys. The watershed algorithm obtains the watershed ridge lines shown in image (d). For ease of comparison, image (e) combines image (b) and (d). Finally, image (f) is obtained by overlaying images (b) with the complement of image (d). The black lines running through some of the kernels are watershed lines that cross an individual kernel.

Inspection of Figures 1(e) and 1(f) shows that many corn grains appear oversegmented, due to the fact that many corn kernels have more than one local maximum. Figure 2(a) shows the corn seed contours along with the local maxima obtained from the distance-transformed image in Figure 1(c). Figure 2(b) is an enlarged section of the same image clearly showing multiple local maxima on several seeds.

Comparison of Figures 2 and 1(e) clearly shows that there is no unique local maximum for some corn kernels based only on the distance transform, leading directly to the oversegmentation.

2.2. Extended-Maxima Transform. The H-maxima transform in mathematical morphology [17], used to suppress pixels above a certain intensity, can extract local maxima related to target objects from the gray image. The extended-maxima transform can calculate the regional maximum, and constitute a binary image. The H-maxima transformation suppresses all maxima whose depth is below or equal to a given threshold h. This is achieved by performing the reconstruction by dilation of f from $f - h$, as shown in the following:

$$H\,\mathrm{MAX}_h(f) = R_f^{\delta}(f - h), \tag{1}$$



(a) (b)

Figure 2: (a) Local maxima superimposed onto a contour image. (b) Local enlarged image.

163	168	188	179	112	14	17	12	12	18	25	25
175	179	181	181	162	36	14	23	24	15	11	17
179	186	198	197	180	59	10	16	20	21	23	26
184	200	199	196	192	55	7	22	23	16	21	25
192	196	195	207	179	38	15	21	22	25	27	22
197	209	196	201	142	16	20	24	23	23	27	27
197	207	203	209	91	7	20	26	27	23	25	29
211	200	205	176	34	19	22	14	13	22	27	26
198	204	206	96	6	19	47	55	65	57	37	25
192	188	165	91	139	191	223	231	235	232	220	196

(a)

163	168	171	171	112	14	17	12	12	18	23	23
171	171	171	171	162	36	14	21	21	15	11	17
171	171	171	171	171	59	10	16	20	21	23	25
171	171	171	171	171	55	7	22	23	16	21	25
171	171	171	171	171	38	15	21	22	25	27	22
171	171	171	171	142	16	20	23	23	23	27	27
171	171	171	171	91	7	20	23	23	23	25	27
171	171	171	171	34	19	22	14	13	22	27	26
171	171	171	96	6	19	47	55	65	57	37	25
171	171	165	91	139	172	172	172	172	172	172	172

(b)

0	0	1	1	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0	0
1	1	0	0	0	1	1	1	1	1	1	1

(c)

Figure 3: (a) Original matrix, (b) H-maxima transform, and (c) extended-maxima transform.

where f is the grayscale of original grayscale image, and h is the threshold value. The extended-maxima EMAX are defined as the regional minima of the corresponding H-maxima transformation, as shown in the following:

$$E\,\mathrm{MAX}_h(f) = R\,\mathrm{MAX}\left[H\,\mathrm{MAX}_h(f)\right]. \tag{2}$$

The previous transformations can be described by Figure 3. The images are the regions with coordinates from (115, 86) to (126, 95) based on Figure 1(a), and the threshold value here is set to $h = 170$.

2.3. Watershed Segmentation Algorithm Based on the Extended-Maxima Transform. In this paper, the extended-maxima transform is used to eliminate oversegmentation. The algorithm is as follows: first, we perform a distance transformation on the original grayscale image. Next, we need to look for an appropriate value for the threshold parameter h; then, the H-maxima transformation and extended-maxima transformation are applied sequentially to the distance-transformed grayscale image. After these transformations, we obtain a binary image. Watershed transform on this binary image yields watershed ridge lines. These ridge lines are then superimposed on the original binary image. Experimental data show that touching corn kernels can be segmented well by these watershed ridge lines, provided that h is selected properly.

The key point of the algorithm is how to select the optimal threshold value h, which can segment the original binary image correctly. In order to improve the calculation precision, firstly, the transformed image is normalized. That means that the grayscale will be in the range of $[0, 1]$ after the distance transformation. We then repeat the entire algorithm after increasing h by an increment Δh until we obtain the optimum value for h. If h is greater than the optimum value range, it will not eliminate the oversegmentation. However, if h is less than this range, "segmentation debris" (spurious bright or dark pixels that were not cleared out by the processing) will appear after segmentation, making subsequent image analysis processing unnecessarily difficult.

This paper uses the connected component labeling method [18] to count the region number of kernel images and determine the optimal range of the threshold value h. Firstly, the number of kernels in the binary image created by the extended-maxima transform is calculated, and the results are recorded as N_1. The watershed segmentation ridge lines based on the extended-maxima transform are superimposed to the original binary image; the segments results are recorded and counted as N_2. When an allowable range of the threshold parameter h is decreased, kernels in the binary image from the extended-maxima transform can be separated because the kernel edges are brighter than the main kernel. Unlike erosion operations, the extended-maxima transform can avoid discarding small objects as spurious, so it will not

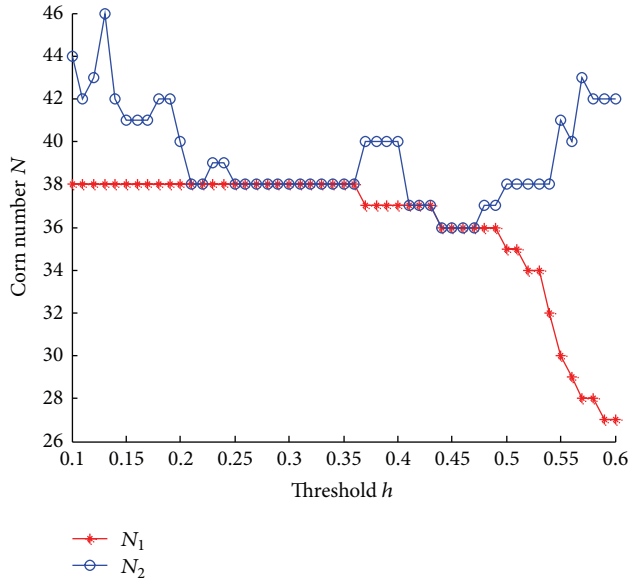


FIGURE 4: N_1-N_2 statistical data curve.

lose small particles through improper selection of structural elements in the erosion operations. Experimental data show that it can accurately predict the number of kernels when a small enough threshold h is selected. Taking the collection of corn kernels in Figure 1 as an example, setting different values of h gives different values of both N_1 and N_2 as shown in Figure 4. The data represent a stepwise change to N_1 with the increase of the threshold h. When threshold h is small, N_1 is constant and equal to the actual numbers of kernels in the image. We choose as the selection criterion the first value of h where N_1 is equal to N_2.

In conclusion, the steps of extended-maxima transform/watershed separation algorithm are as follows.

(1) Convert the original grayscale image to binary and distance transform it. Then, normalize the obtained grayscale image. We call the resulting image I_{bw}.

(2) Guess an initial threshold value $h = 0.1$, and perform an extended-maxima transform on I_{bw}. We call the new binary image I_{bw1}.

(3) Count the number of kernels in I_{bw}, which is based on connected components labeling; statistical number N_1 is obtained.

(4) From a watershed segmentation transform on I_{bw1}, we obtain watershed ridge lines and superimpose them onto the original binary image, obtaining a segmented image with each segment corresponding to a single kernel.

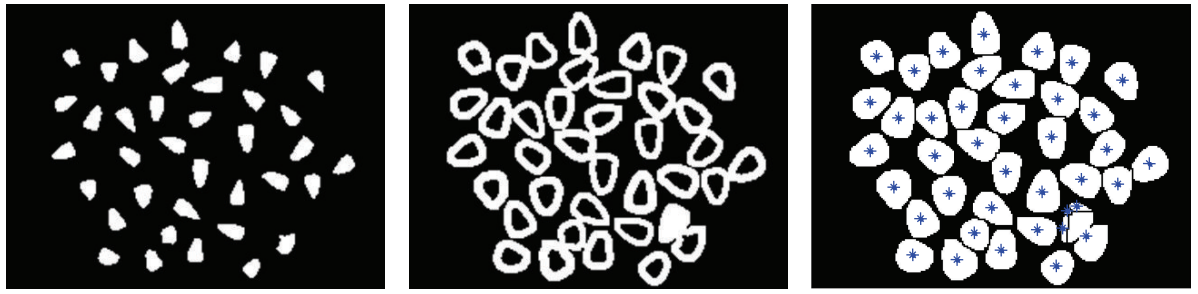
(5) Count the number of kernels in segmented image, which is based on connected components labeling; statistical number N_2 is obtained.

(6) If $N_1 \neq N_2$, $h = h + \Delta h$ (here take $\Delta h = 0.01$), repeat steps (2), (4), (5), and (6), until $N_1 = N_2$.

3. Experiment and Analysis

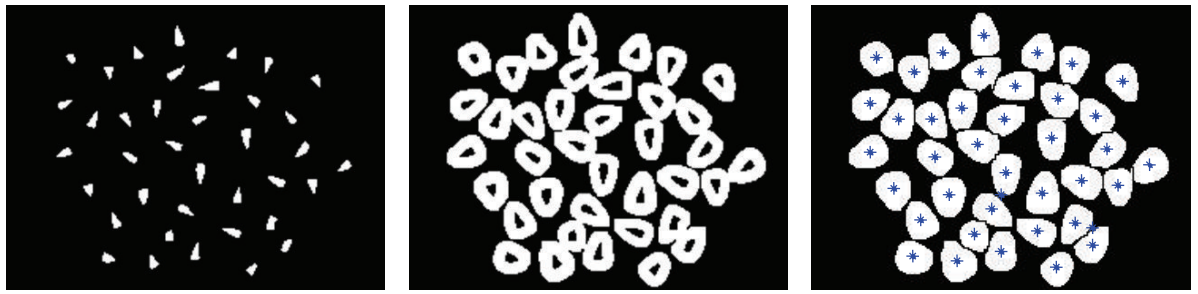
3.1. Oversegmentation and "Segmentation Debris". An extended-maxima transform and watershed segment transform are performed on the original image (Figure 1(a)) using different threshold values $h = 0.4$ and $h = 0.2$. For ease of comparison, the binary image based on the extended-maxima transform is subtracted from the original binary image, and the centroid of the connected regions (kernels) are marked with "$*$." When $h = 0.4$, oversegmentation occurs with the result $N_1 = 37$, $N_2 = 40$, as shown in Figure 5. On the other hand, when $h = 0.2$, there is some "segmentation debris" (the small particles cutting from the kernel region by watershed segmented ridge lines), and the result $N_1 = 38$, $N_2 = 40$ is shown in Figure 6.

By analyzing Figures 5 and 6, we conclude that the value of the threshold parameter h has great influence on the segmentation effect. When h is too large, as shown in Figure 5, touching kernels cannot be effectively segmented. Figure 5(b) shows the result of subtracting the transformed image from the original binary image. The extended-maxima transform result as shown in Figure 5(a) is similar to the original image, but with less white space per kernel. It is this "similar scaling" that enables the extended-maxima transform watershed segmentation algorithm to correctly segment touching seeds. Due to the irregularity of the size and shape of corn kernels, the higher the similarity between extended-maxima transformed image and original binary image is, the more likely some kernels have been lost. Figure 5(c) is an example of oversegmentation. However, when h is relatively small, Figure 6(b) shows that although every single kernel on the extended-maxima transformed binary image has shrunk the small kernels do not reflect the outlines of the original binary image very well. After the watershed segmentation, the segmented image shows "segmentation debris" at the border of touching kernels, shown in Figure 6(c). The number of



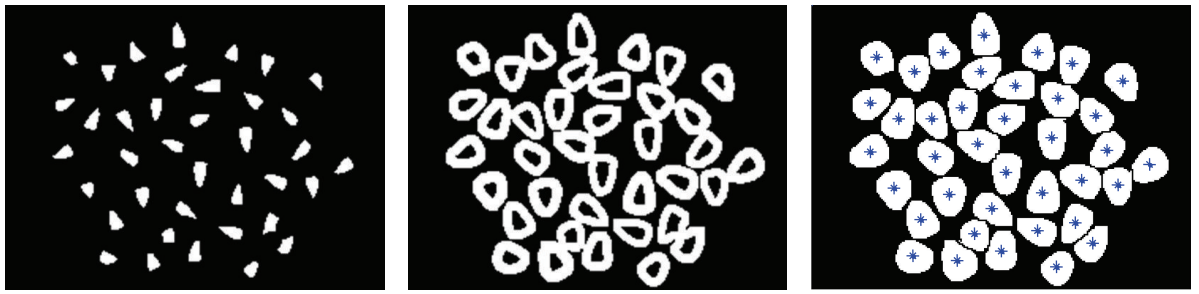
(a) (b) (c)

Figure 5: Oversegmentation (a) extended-maxima transform, (b) subtraction, and (c) labeled segments.



(a) (b) (c)

Figure 6: Segmentation debris (a) extended-maxima transform, (b) subtraction, and (c) labeled segments.



(a) (b) (c)

Figure 7: Correct segregation (a) extended-maxima transform, (b) subtraction, and (c) labeled segments.

kernels obtained by this process is often greater than the real number, but "segmentation debris" is often difficult to observe, so the miscount can easily go unnoticed.

3.2. Correct Segmentation. Observation and analysis of Figure 4: within the range of h before N_1 has undergone its first step change, the points on N_2 curve intersect the N_1 curve giving the optimized segmentation threshold value h. The optimal range of the threshold parameter h can be obtained from Figure 4; in this case, there are two ranges, namely (0.21, 0.22) and (0.26, 0.36). Correct segmentation results are obtained in these threshold ranges. For example, taking h = 0.3 gives the extended-maxima transform image, subtraction image, and segmented labeled image shown in Figure 7.

3.3. Experiment Results. We calculate the accuracy of the algorithm by segmenting 50 images, each of which contained 400 corn kernels. The results are shown in Table 1. The average number of segments is 399.46, the standard deviation is 1.05 segments, and the average segmentation accuracy rate is as high as 99.87%.

To observe the smoothness of the segmentation curve, we get boundaries of segmented regions which are labeling separately by different color, as is shown in Figure 8. By observing and analyzing the boundaries of the labeling region below, what can be concluded is that the majority of split curve is smooth between the adhesion of the particles. As for the closer and complicated adhesion phenomenon for some kernels, it can be seen from the segmented figure that few split curves show distortion phenomenon.

Table 1: The segmentation results of touching corn kernels.

h	N_1/N_2	Segmentation number	Accuracy rate
0.14	399/400	397	99.25%
0.12	400/400	400	100.00%
0.15	400/400	400	100.00%
0.17	400/400	400	100.00%
0.10	400/400	400	100.00%
0.13	400/400	400	100.00%
0.12	400/401	399	99.75%
0.14	400/402	398	99.50%
0.12	400/400	400	100.00%
0.12	400/400	400	100.00%
0.12	400/400	400	100.00%
0.10	400/401	399	99.75%
0.15	400/401	399	99.75%
0.16	400/400	400	100.00%
0.17	400/400	400	100.00%
0.14	400/402	398	99.50%
0.16	400/400	400	100.00%
0.18	400/400	400	100.00%
0.16	400/400	400	100.00%
0.16	400/400	400	100.00%
0.10	400/402	398	99.50%
0.18	400/400	400	100.00%
0.18	400/400	400	100.00%
0.15	400/400	400	100.00%
0.19	400/400	400	100.00%
0.14	400/400	400	100.00%
0.21	400/400	400	100.00%
0.22	400/400	400	100.00%
0.16	400/400	400	100.00%
0.17	400/401	399	99.75%
0.18	400/400	400	100.00%
0.10	400/401	399	99.75%
0.12	400/400	400	100.00%
0.16	400/400	400	100.00%
0.13	400/400	400	100.00%
0.24	400/400	400	100.00%
0.15	400/400	400	100.00%
0.17	400/400	400	100.00%
0.11	400/401	399	99.75%
0.13	400/400	400	100.00%
0.20	400/400	400	100.00%
0.20	400/400	400	100.00%
0.10	400/400	400	100.00%
0.13	400/400	400	100.00%
0.11	400/401	399	99.75%
0.12	400/403	397	99.25%
0.12	400/400	400	100.00%
0.12	399/405	395	98.75%
0.17	400/400	400	100.00%
0.16	400/402	398	99.50%

4. Conclusion

From the previous theoretical analysis and experimental results, we have shown that irregularly shaped corn kernels usually are oversegmented by the traditional watershed segmentation algorithm, which is harmful to subsequent image processing. This paper proposes an improved watershed algorithm based on the extended-maxima transform in which the threshold value parameter h is gradually increased. By comparing the statistical number of connected regions between an image transformed with the extended-maxima transform and that transformed with the watershed segment algorithm, the optimal range of threshold value can be obtained. Experimental data shows that, within the optimal range of the threshold value, touching kernels can be effectively segmented by the proposed algorithm, and the correct number of kernels can be obtained. The accuracy of segmentation is as high as 99.87%.

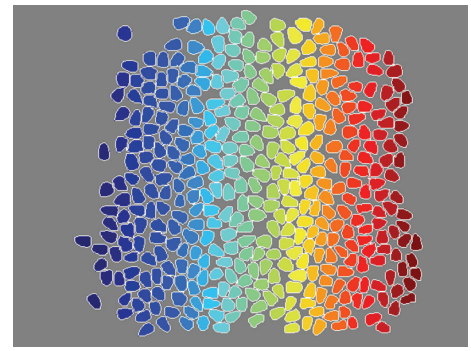


Figure 8: Boundaries of segmented region.

Acknowledgments

The authors gratefully acknowledge assistance from Professor Mike Hey from the University of Nottingham, Dr. John S. Hyatt from the Georgia Institute of Technology, and also give thanks to the China National Science and Technology Support Program (Grant no. 2012BAK08B04) and the Natural Science Foundation of Shandong Province (Grant no. ZR2012FL22).

References

[1] Z. Yaqiu, W. Wenfu, and W. Gang, "Separation of corn seeds images based on threshold changed gradually," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 27, no. 7, pp. 200–204, 2011 (Chinese).

[2] P. Shatadal, D. S. Jayas, and N. R. Bulley, "Digital image analysis for software separation and classification of touching grains: I. Disconnect algorithm," *Transactions of the American Society of Agricultural Engineers*, vol. 38, no. 2, pp. 635–643, 1995.

[3] W. Wang and J. Paliwal, "Separation and identification of touching kernels and dockage components in digital images," *Canadian Biosystems Engineering*, vol. 48, pp. 1–7, 2006.

[4] L. Yan, C.-W. Park, S.-R. Lee, and C.-Y. Lee, "New separation algorithm for touching grain kernels based on contour segments and ellipse fitting," *Journal of Zhejiang University Science C*, vol. 12, no. 1, pp. 54–61, 2010.

[5] Y. Xun, G. Bao, Q. Yang, F. Gao, and W. Li, "Automatic segmentation of touching corn kernels in digital image," *Transactions of the Chinese Society of Agricultural Machinery*, vol. 41, no. 4, pp. 163–167, 2010 (Chinese).

[6] N. S. Visen, N. S. Shashidhar, J. Paliwal, and D. S. Jayas, "Identification and segmentation of occluding groups of grain kernels in a grain sample image," *Journal of Agricultural Engineering Research*, vol. 79, no. 2, pp. 159–166, 2001.

[7] G. Hua, W. Yaqin, and G. Pingju, "Research on segmentation algorithm of adhesive plant grain image," in *Proceedings of the 8th International Conference on Electronic Measurement and Instruments (ICEMI '07)*, pp. 2927–2930, August 2007.

[8] G. Zhang, D. S. Jayas, and N. D. G. White, "Separation of touching grain kernels in an image by ellipse fitting algorithm," *Biosystems Engineering*, vol. 92, no. 2, pp. 135–142, 2005.

[9] G. Zhang, J. Deng, and D. S. Jayas, "Separation touching grain kernel in machine vision application with Hough Transform and morphological transform," in *Proceedings of the Canadian Society of Agricultural Engineering (CSAE/SCGR '05)*, Paper no. 05-006, June 2005.

[10] W. Wang and J. Paliwal, "Separation and identification of touching kernels and dockage components in digital images," *Canadian Biosystems Engineering*, vol. 48, pp. 7.1–7.7, 2006.

[11] E. H. Van den Berg, A. G. C. A. Meesters, J. A. M. Kenter, and W. Schlager, "Automated separation of touching grains in digital images of thin sections," *Computers and Geosciences*, vol. 28, no. 2, pp. 179–190, 2002.

[12] Y.-C. Wang and J.-J. Chou, "Automatic segmentation of touching rice kernels with an active contour model," *Transactions of the American Society of Agricultural Engineers*, vol. 47, no. 5, pp. 1803–1811, 2004.

[13] R. C. Gonzalez, *Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ, USA, 2002.

[14] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, 1991.

[15] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," *Journal of the ACM*, vol. 13, no. 4, pp. 471–494, 1966.

[16] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer, New York, NY, USA, 2003.

[17] S. Halkiotis, T. Botsis, and M. Rangoussi, "Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks," *Signal Processing*, vol. 87, no. 7, pp. 1559–1568, 2007.

[18] K. Suzuki, I. Horiba, and N. Sugie, "Linear-time connected-component labeling based on sequential local operations," *Computer Vision and Image Understanding*, vol. 89, no. 1, pp. 1–23, 2003.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 805363, 10 pages
http://dx.doi.org/10.1155/2013/805363

Research Article

Research on Optimization for Motion Control Bus Based on Ethernet

Kai Sun,[1] Jie Zhang,[1] Guangping He,[1] and Youdong Chen[2]

[1] *Department of Mechanical and Electrical Engineering, North China University of Technology, Beijing 100144, China*
[2] *School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Kai Sun; sunkai@ncut.edu.cn

Received 8 April 2013; Accepted 13 May 2013

Academic Editor: Yong Tao

Field bus system has been successfully introduced into industrial automation. Nowadays, most of the motion control bus is based on the Ethernet physical layer, and all of the new standards are based on the Ethernet physical layer. This paper introduces a new optimized technology based on motion control bus of the Ethernet physical layer, which includes the prediction mechanism in the arrival of the frames, the retransmission mechanism in advance, and a new mechanism of independent verification of data segment. Tests show that using the mechanisms in this paper can enhance efficiency and reliability in communication process.

1. Introduction

In recent years, international advanced NC machine and control system of industrial robot all have used motion control bus to connect controller, servo and I/O equipment, and so forth. Motion control system based on motion control bus has advantages of freely amending hardware configuration, easily achieving distributed control, and reliability and higher control system accuracy.

Recently, most of the motion control bus is based on RS485 and Ethernet physical layer, and all of the new standards are based on Ethernet physical layer; typical examples are from MECHATROLINK II to MECHATROLINK-III, from SERCOS II to SERCOS III, and from CC-Link to CC-Link IE, all of them have changed into Ethernet physical layer from the RS485 physical layer of the last generation; it is a common sense to use Ethernet physical layer as a new generation of physical layer of motion control bus. NC machines and control system of industrial robots have different control requirements in the field bus contrasted by ordinary control system. Generally speaking, the motion control system has requirements for synchronization and real time that is, to say, the synchronous cycle is at least 1 ms, and synchronous jitter is at least 1 μs [1]. But original function of Ethernet is not designed for the field bus, so different field bus standards generally take some transformations which are based on typical Ethernet. From the view of means of transformation, this can basically be divided into two categories

(1) *Pure software type*. It is characterized by increasing real time [2–5] through changing network protocol stack [6], system software of slave station, or scheduling method of switch. This method does not need to change existing hardware and has become one of hot spots [7, 8] because of the lowest cost and the fastest implement, but its synchronization performance and real-time performance are at the worst. In the academic research literature, synchronous cycle of pure software is about 1 ms or so, and in the different open industry standard, synchronization cycle is always 5 ms [9–11] or so. Among them, the best real-time performance is Ethernet Powerlink; it uses a concentrator rather than a switch to avoid delay in the exchange process; that is to say, by using bus topology, it avoids delaying problem of star structure switch, making synchronization cycle up to 200 μs and synchronization jitter up to 1 μs [12], but its bus structures also limit transmission efficiency and performance of synchronization and real-time decreases along with the increase in the number of slave stations [13].

(2) *Soft and hard combination type*. It is characterized by changing MAC layer of slaves station and using the special ASIC chip or FPGA to accelerate its receiver and transmission speed of slave station; the typical method includes EtherCAT,

SERCOSIII, Profinet IRT, MECHATROLINK-III, and RTEX. All of their synchronization cycle can reach less than 100 μs and synchronization jitter can reach less than 1 μs.

Except for the two above-mentioned transformation methods, in recent years, popular IEEE1588 standard [14] can also provide rather good performance of synchronous jitter.

Seen from the comparison of each the above-mentioned of technical routes, the method of pure software type has poor performance and cannot meet the requirements in the real time, so if we use the EtherCAT and the chip SERCOS III based on both ordinary Ethernet physical layer and transform solutions of the link layer, it will be the most suitable in CNC machine and industrial robot control system.

The research on program of transformation of the link layer is being widely and deeply carried out in foreign countries. In 2007, Jasperneite et al. carried out a detailed analysis and comparison of performance and potential about EtherCAT and Profinet IRT [15]. In 2008, Prytz made a detailed comparison between synchronization performance and real-time performance about EtherCAT and Profinet IRT [16]. At the same year, Schumacher et al. made some intensive research on Profinet IRT, putting forward IRT+ protocol based on mode of IRT; his optimized methods included optimization of frame preamble and optimization of different data organization mode from upstream data and downstream data, which was of great value [17]. In 2010, Cena et al. made an analysis for real-time performance [18]. At the same year, Gunzinger et al. made a research on optimization in real-time performance of Profinet IRT and achieved it by using FPGA [19]. From 2010 to 2011, Knezic et al. made an optimization on upper layer protocol of EtherCAT [20] and made an analysis on network topology of EtherCAT [21].

2. Research on Synchronization

Realization mechanism of the IEEE1588 is complex, and its cost is much high, although its synchronous jitter performance (8 ns) is much better than that of the EtherCAT (20 ns), but because of the increase in relative cost, the significance of improvement in this performance is limited for the motion control system. For the ordinary motion control systems, requirements can be met in 8 ns synchronous jitter of SERCOS III, but realization process of the EtherCAT mechanism is still too complex; the function from complex clock system of the EtherCAT mechanism is limited in NC machines and industrial robots system.

Timing completion of calculation results are determined by characteristics of data sampling interpolation methods, and timing repeat is also determined in the process of communication, so synchronous mechanism of SERCOS III depending on the data frame that is timed and sent is very appropriate here. But because of uncertainty of delay in the processing of data received by the Ethernet physical layer chip, synchronization mechanism of SERCOS III in slave station is much more difficult to reliably work at most times.

2.1. Analysis of Crucial Question of Synchronization Mechanism. Uncertainty cannot be avoided in receiving delay process, while ordinary Ethernet PHY chip handles Ethernet frames. This reason caused by uncertainty differs from different work mode of PHY chip. The work mode includes standard MII mode and RMII mode. Under MII's mode and RMII's mode, although delay fluctuation appears in different positions, the results are the same, and this delay fluctuation will constantly expand along with the Ethernet frames spreading in the slave station. By finding of actual measurement and simulation calculation of delay transmission, delay fluctuation is in line with binomial distribution. According to analysis of the simulation results, delay fluctuation of the farthest slave station will continually expand along with increase in network and slave station number, which has serious impact on synchronous ability of the whole network.

2.2. Mechanism for the Arrival of Prediction. In order to solve cumulative problem in delay fluctuations along with the increase in sequence number of slave station, the prediction mechanism in the arrival of frames is used in this paper to revise start time of timer, when delay fluctuation is larger, which make timer start at reasonable time and obtain more stable synchronous signal.

In order to predict the moment of the arrival of frames, firstly we need to get ideal time at the arrival of communicating frames, and then use this moment as time origin and calculate the arrival time of subsequent frames by communication cycle.

The process in the implement in prediction of the arrival of slave station is shown in Figure 1 as follows.

(1) First communication cycle clocked will be told to slave station by master station, this communication stage set is $t1$, then subsequent and respective communication stages are $t2$, $t3$, $t4$,....

(2) When slave station receives communication cycle at stage $t1$, counter C_p will start to count.

(3) Another counter C_f is started by slave station in stage $t2$ and starts to count at the same time. When frames are actually received, C_p is recorded as C_{px}, counter value transformed by communication cycle is recorded as C_t, the difference of C_{px} and C_t is recorded as C_pt, and counter value arrival predicted in next time is recorded as C_o and its value set as $C_{pt.\lim}$. From the start of communication stage between $t2$ and $t2$ later, the interval time to receive the next frame is predicted when frame is received every time; the calculation method is shown as follows: if C_{pt} is larger than $C_{pt.\lim}$, then $(C_t + C_{pt.\lim})$ will be decreased when C_p is counting at the same time, and C_o is set as $(C_t + C_{pt.\lim})$; if C_{pt} is smaller than $C_{pt.\lim}$, then $(C_t - C_{pt.\lim})$ will be decreased when C_p is counting at the same time, and C_o is set as $(C_t - C_{pt.\lim})$ if the range of C_{pt} is from negative $C_{pt.\lim}$ and positive $C_{pt.\lim}$, C_t will be decreased when C_p is counting at the same time, and C_o is set as C_t.

(4) Within stage $t3$ and its later stage, when count value C_f reaches value C_o obtained by the calculation of the upper stage, signal will be reached by frames which generate prediction, and at the same time the frames themselves will be cleared to start counting from 0. That arrival signal obtained by prediction from frames by every time is considered as arrival signal from actual frames will start timer that generates synchronous signal and will generate synchronous

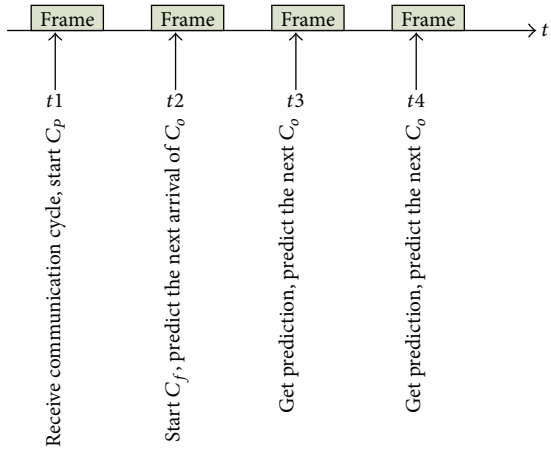


Figure 1: Timing diagram of predicting the arrival mechanism.

signal as SERCOS II's mechanism of generating synchronous signal. Because of the situation that there are two timers in SERCOS II's mechanism, slave station needs to provide a forecast system for every timer.

In the above-mentioned prediction of the arrival moment of frames, the smaller the value of $C_{pt.\lim}$, the better it will be, but the condition must be met as follows:

$$\frac{C_{pt.\lim}}{F_s} \geq T \times S_r. \tag{1}$$

Thereinto, F_s is clock cycle of slave station, and S_r is relative drift difference value for clock between master station and slave station. If $C_{pt.\lim}$ degenerates to 0, the interval of predicted arrival moment from slave station will become C_t, losing compensation function for the clock drift. Because of the situation that $C_{pt.\lim}$ cannot be decimal, minimum of $C_{pt.\lim}$ will be less than 1 with the reduction of T, so finally in case of the situation that the value will be less than 1, value of $C_{pt.\lim}$ will be 1.

Figure 2(a) is fluctuation simulation waveforms of actual arrival interval frames of the twenty-fourth slave station when F_s is 50 MHz, and Figure 2(b) is fluctuation simulation waveforms of interval of predicted arrival moment of the station. Fluctuations of 480 ns can be reduced to the range of 160 ns by using testing mechanism at the same time jitter of predicted arrival moment that is only 160 ns.

According to the simulation results of slave station which ranges from 1 to 48, when F_s is 50 MHz, jitter of actual arrival moment of four previous slave stations is within 160 ns; jitter will begin to constantly increase from the fifth slave station, and jitter of the predicted arrival moment can be controlled within 160 ns all the way.

There is a small probability that the interval of actual arrival moment of frames is more than 400 ns, there is no way to provide comparative effect in prediction mechanism under such situation, but according to the analysis from the view of qualitative point, it is assumed that arrival moment of a frame is severely later, which is closer to the maximum of fluctuation interval, and according to the above-mentioned distribution rules, probability that next frame is similar with the situation is very small, therefore although the interval of the frame that is, severely hysteretic and the previous frame will be very large, but the interval of the frame and the next frame will be very small, which will completely be filtered out by integral low-pass filter. When value of $C_{pt.\lim}$ is 1 and F_s is 100 MHz, even consecutive and serious lag is aroused in frames, which cause that the maximal fluctuation of delay is just $m \times 20$ ns. This extreme situation can be dealt with by integral low-pass filter in the prediction mechanism too.

2.3. The Improvement of the Arrival of Prediction in Violent Interferential Environment. In some severe interference environment, MST telegrams of SERCOS III can also be interfered, and synchronous signal completely is lost by slave station at the same time. By using the above-mentioned mechanism for arrival of prediction in this article, even though the frames weren't completely received by slave station, synchronous signal can also be generated by slave station by means of keeping principle invariable and modifying slightly. The principle is shown as follows:

$$C_{s.\max.n} = \left(T + n \times 20 \times 10^{-9}\right) \times F_s. \tag{2}$$

In the second step of prediction realization process of slave station, a new added counter is used for recording interval of frames, when value of C_s is greater than value of $C_{s.\max}$, it is considered that a frame is completely lost, at the same time, let us assume that a frame is received,and continues to run as original steps. $C_{s.\max}$ is maximum value of the internal of receiving frames in theory, $n \times 40$ ns is the maximum value of the interval fluctuation of receiving frames from the nth slave station, $C_{s.\max.n}$ is $C_{s.\max}$ of the nth slave station, which is shown as follows:

$$C_{s.\max.n} = \left(T + n \times 20 \times 10^{-9}\right) \times F_s. \tag{3}$$

If a frame is completely lost by detection, then when detection is performed again, we need to consider $T \times F_s$ as value of $C_{s.\max.n}$, because there has been judged on maximal lag when frames are lost by judgment in the first time, it is impossible that the maximum of the interval of the first lost frame and the second lost frame is greater than lagging mean $T \times F_s$.

3. Research on Real-Time

3.1. Limit of the Ethernet Protocol for the Transmission of Frames. The transition of straight-through principle can greatly improve efficiency of transmission of ring topology, but due to some of the Ethernet physical layer's requirements and the limitations of the physical layer chip, there is still space for efficient improvement in the field of motion control. Requirements and limitations are shown as follows:

(1) There must be a minimum time interval between two separate Ethernet frames (front and back) in one line, we can call the minimum time interval as minimum interval between frames.

(2) Ethernet frames have limit of minimum length. In order to avoid the problem in collisions, CSMA/CD

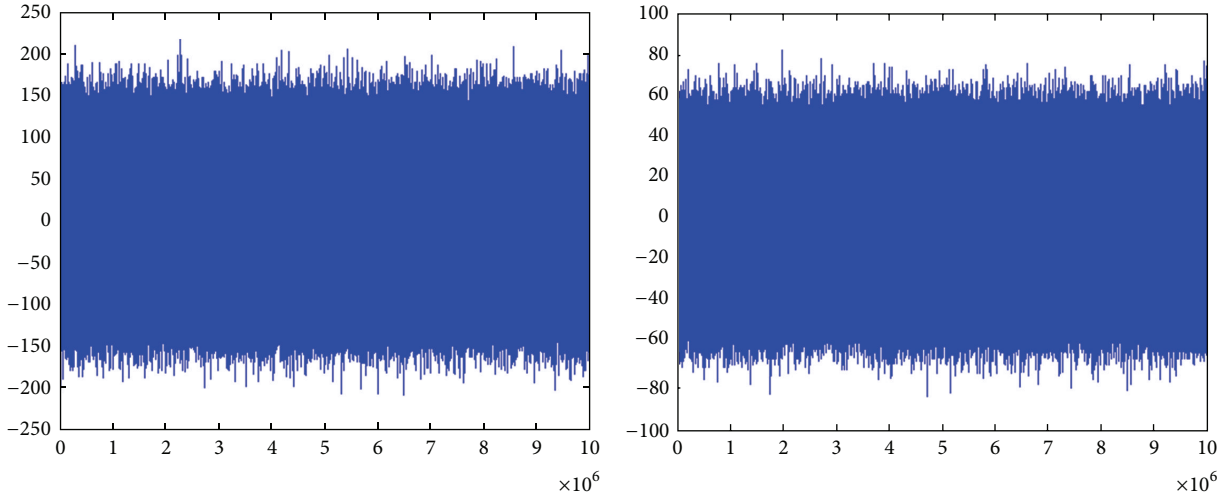


Figure 2: Simulation comparison of arrival prediction mechanism of 24th slave station.

mechanism is used in transmission protocol of standard Ethernet. The minimum length of frames is 64 bytes, which is ruled by transmission protocol of standard Ethernet.

3.2. Retransmission Mechanism in Advance

3.2.1. Contradictions of Real Time and Reliability. As for efficient research in normal communication condition, currently problems have been solved better by using lumped-frame direct forwarding principle, whose efficiency has been closed to limit in theory. From the view of the conditions of actual testing in this paper, as for bus using twisted-pair physical connection, under the conditions of severe electromagnetic interference, it is very common to see communication failure. As for the conditions of communication failure, the conditions can generally be dealt with by retransmission mechanism. As for master station, the paper notes that the network is idle in the period of time between the end of the first frame transmitted and the starting of retransmission of frames, and idle proportion will be increased, while the amount of datum from the slave station is smaller and transmission speed of the Ethernet is higher. The reason which causes larger idle is related to the delay of Ethernet physical layer chip.

3.2.2. Delay of Transceiver from Physical Layer Chip. As for Ethernet of 100 M, for example in NS Company's Ethernet PHY chip DP83848, the average of delay of transceiver is 590 ns by using RMII interface and cable of 0 m between transceiver stations, and average of the same delay is 310 ns by using MII interface and cable of 0 m between transceiver stations, but the signal is required for special processing while clock domains are crossed by it, because RX_CLK of MII mode is not synchronous with local clock, the process of processing itself is about four local clocks, for example, additional 80 ns is needed for local clock of 50 MHz. As for Ethernet of 1000 M, for example in NS Company's Ethernet PHY chip DP83848, according to the datum given by the data sheet, delay of transmission from GMII interface is 152 ns and delay of receiver from that is 386 ns; therefore, delay of transceiver is 536 ns. In conclusion, the set of delay of transceiver for every node is 0.5 μs, which is shown as t_d.

3.2.3. Retransmission Mechanism in Advance. As for the situation of retransmission, since the retransmitted data itself does not depend on contents of the first data frame, frames can be retransmitted in advance in this idle time; that is to say, frames must be retransmitted whether it is necessary or not; the schematic is shown in Figure 3. In this paper, this mechanism is called retransmission mechanism in advance.

When retransmission frames are transmitted by master station in advance and number of retransmission is r, considering the above-mentioned minimum interval in the process of transmitting frames by master station, minimum of communication cycle is obtained as follows (unit is μs):

$$T = \frac{m \times n \times 4}{10^g \times 5} \times (r+1) + n \times t_d \times h + (0.96 + n \times 0.04 \times h) \times r. \quad (4)$$

After using retransmission mechanism in advance, delay of communication cycle can be significantly reduced, which is caused by retransmitting frames. Take an example of the above-mentioned conditions.

(1) When datum from slave station is 8 bytes, number of slave station will be 24, by using single ring and Ethernet of 100 M, and number of retransmission is 1, transceiver time of each frame of master station is 39.36 μs, communication cycle is 78.72 μs, and transceiver time of retransmitting frames is also 39.36 μs. After using retransmission in advance, time is occupied by retransmission part for about 18.24 μs (15.36 μs + 2.88 μs), communication cycle will be 57.6 μs, and time will decrease to 26.8%.

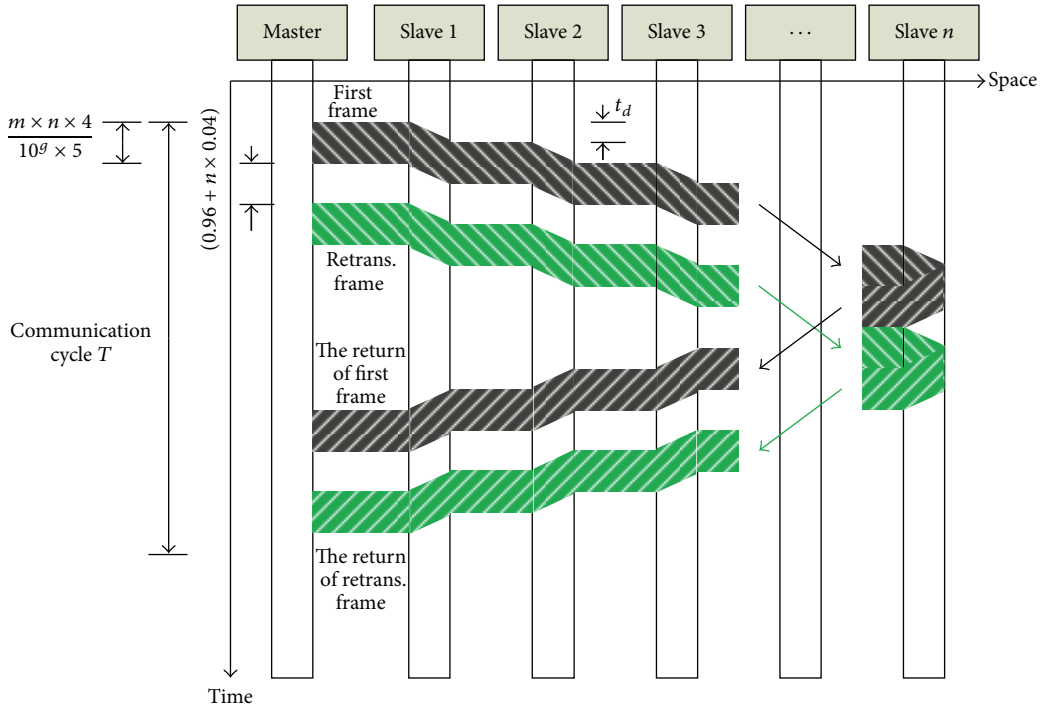


Figure 3: Diagram of data frame of advance retransmission mechanism in the process of transmission.

(2) If by switching to use Ethernet of 1000 M the transceiver time of each frame is about 25.54 μs, the communication cycle which is added time-consuming is 51.07 μs. After using the retransmission in advance, time that is occupied by the part in retransmission is about 4.42 μs (1.54 μs + 2.88 μs), the communication cycle will be 29.95 μs, and the time will decrease to 41.2%.

As for using the Ethernet of 1000 M, from another point of view, After using retransmission mechanism, communication cycle (29.95 μs) is only more 4.42 μs than the same situation without using the retransmission mechanism, and extra parts account for 17.3%, that is to say, the increase in reliability can multiply be got at the cost of the increase 17.3% in communication cycle. Therefore, retransmission mechanism in advance is particularly suitable for utilization in large-scale and high-speed network.

4. Research on Reliability

Software mechanism of improving bus reliability is mainly about retransmission mechanism. The retransmission mechanism is a software mechanism, which has a low cost and advantages of high reliability, but the retransmission is at the expense of the reduction in communication cycle.

4.1. Contradiction of Reliability and Efficiency. For issues of minimum interval between data frames and frame leading, communication efficiency of bus can be effectively improved by using the lumped frame instead of multiple single-node frames, when the single-node data is less. When the interference is very frequent and the interference cycle is shorter than communication cycle of bus, the retransmission mechanism will lose its original significance.

The authors note that the existing different standards about data format of slave station is not specially designed for how to improve the reliability under the condition of retransmission, when the retransmission mechanism is used, thus relationship between the degree of improvement in reliability and number of retransmission is still not ideal. So an independent testing mechanism in data segments from lumped frame is proposed, and reliability in communication in the condition of retransmission can be significantly improved by using this mechanism.

4.2. Independent Verification of Data Segment. The system reliability can be improved by mechanism of independent verification of data segment from the perspective of reliability model, which can be proved as follows:

$$R_{ps} = 1 - \prod_{i=1}^{m} \left(1 - \prod_{j=1}^{n} r_{ij}\right). \tag{5}$$

Figure 4 is a block diagram of reliability of parallel-serial system, whose mathematical model of reliability is shown as follows.

R_{ps} is the reliability of the overall system, r_{ij} is the reliability in the jth data segment from the ith intraframe, m is the number of communications, and n is number of slave stations. When the random interference and the length of the data segment from slave station are all the same, all of r_{ij} are the same, and all of r_{ij} values are set as r, and (5) can be simplified as follows:

$$R_{ps} = 1 - (1 - r^n)^m. \tag{6}$$

The reliability of the system model can be described by using typical string and system; Figure 5 is a block diagram

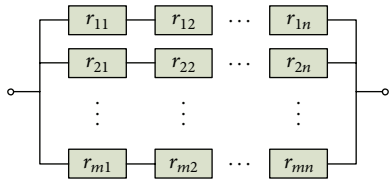


Figure 4: Block diagram of reliability in parallel and series system.

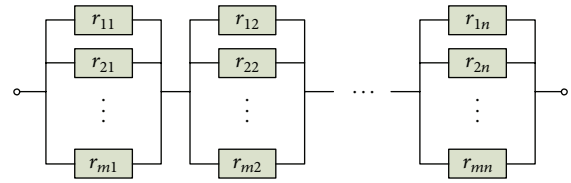


Figure 5: Block diagram of reliability in series and parallel system.

of reliability in series and parallel system, the mathematical model of reliability is shown as follows:

$$R_{sp} = \prod_{j=1}^{n} \left[1 - \prod_{i=1}^{m} \left(1 - r_{ij}\right)\right]. \quad (7)$$

When random interferences and the lengths of the data segment from slave station are the same, all of r_{ij} are the same, and all of r_{ij} values are set as r, and (7) can be simplified as follows:

$$R_{sp} = \left[1 - (1 - r)^m\right]^n. \quad (8)$$

Considering general rule of the reliability, under the condition of the same in unit reliability, the reliability of series and parallel system is higher than the reliability of parallel and series system; therefore, reliability advantages of independent verification mechanism of data segment proposed by this paper can be proved.

In the process of specific implementation in independent verification mechanism of data segment, due to the need for adding parity data to the existing data segment, thus the overall efficiency of the bus transmission is decreased. When 16-bit CRC is used by verification of data segment, the example in this paper is comparatively analyzed as follows:

(1) When slave station data are 8 bytes, numbers of slave station are 24, and by using single ring and Ethernet of 100 M, two-byte CRC verification is added in data from each slave station; transceiver time for each frame of master station increases from 43.2 μs to 48 μs, communication cycle is 96 μs, and time is prolonged by 11.1%.

(2) If we change Ethernet of 1000 M to use without using retransmission in advance, transceiver time for each frame increases from 25.9 μs to 26.4 μs, communication cycle is 52.8 μs, and time is prolonged by 1.8%.

Something still needs to be noted in the process of specific implementation in independent verification mechanism of data segment; in the protocols such as EtherCAT, the data lengths of each slave station are stored within the data segment; when data are searched by slave station, it is needed to use this length to determine the position of next data segment.

4.3. Changing Mechanism of Data Segment Address. As for some extreme cases, interference cycle may be exactly the same with communication cycle; at the same time, whether segmented CRC verification mechanism is used by lumped frame or not may cause master station communicate with one or few slave stations.

As for this extreme case, sending frame time can be determined by characteristics of interference cycle which can be chosen by master station, except that reliability can be improved by autochange in position of data segment in lumped frames, which is called changing mechanism of data segment address in this paper.

Without using independent verification mechanism of data segment, no matter how to change position of data segment in intraframe, all of the data will be lost for the error of FCS in the all frames. By using independent verification mechanism of data segment, the data segment cannot rely on the FCS of entire frames and independently judges whether the data is correct or not; therefore, position change of data segment in frame can help some slave stations avoid interference so as to achieve the purpose of improvement in system reliability.

5. Verification Test and Design for Motion Control Bus

Firstly, this section has designed a set of basic communication protocol in motion control bus the optimization mechanism proposed in this paper is used in the protocol, and then we have used verification platform to achieve this protocol and proved effectiveness of the mechanism proposed in the paper by the actual operation and reliability verification.

5.1. Low-Level Protocol in Bus. In order to verify the optimization mechanism proposed in this paper for motion control bus, a set of basic protocol based on 100 M Ethernet physical layer for motion control bus was designed and implemented in the paper. Frame format of the protocol is shown in Figure 6, and the corresponding field is described in Table 1.

Realization principle for independent verification mechanism of data segment is shown as follows: Sn_LENGTH defines the length of the data segment of the nth slave, slave station x can determines the location of its own data segment according to accumulated values from S1_LENGTH to S (x – 1)_LENGTH, both INDEXLENGTH and INDEX from index segments have verification ICS to can ensure the reliability of the index. Precondition for realizing mechanism of the arrival prediction is shown as follows: the protocol in this paper requires frames timely sent for master station to be transmission interval is sent to slave station through data segment of slave station after completion of initialization for address of slave station. To simplify the design, the protocol requires that a frame can be considered to receive after only detecting correct ICS. The retransmission mechanism in advance is relatively simple to achieve. In order to ensure normal work for the arrival of prediction mechanism and

Table 1: Description for frame format of motion control bus.

4	FCS	CRC verification of standard Ethernet, the verification ranges from LENGTH to DATA
1	Sn_LENGTH	Data length of the nth slave station is length from CMD/STA to SCS, maximum is 255
Determination by Sn_LENGTH	Sn_DATA	Data segment of the nth slave station
2	CMD/STA	Master command or master state
Determination by Sn_LENGTH	DATA0-DATAn	Master data/slave data
2	SCS	Verification of data segment of slave station, verification ranges from CMD/STA to DATAn, CRC16, polynomial is $1 + x^5 + x^{12} + x^{16}$
4	FCS	CRC verification of standard Ethernet, the verification ranges from LENGTH to DATA
1	Sn_LENGTH	Data length of the nth slave station is length from CMD/STA to SCS, maximum is 255
Determination by Sn_LENGTH	Sn_DATA	Data segment of the nth slave station
2	CMD/STA	Master command or master state
Determination by Sn_LENGTH	DATA0-DATAn	Master data/slave data
2	SCS	Verification of data segment of slave station, verification ranges from CMD/STA to DATAn, CRC16, polynomial is $1 + x^5 + x^{12} + x^{16}$
4	FCS	CRC verification of standard Ethernet, the verification ranges from LENGTH to DATA
1	Sn_LENGTH	Data length of the nth slave station is length from CMD/STA to SCS, maximum is 255
Determination by Sn_LENGTH	Sn_DATA	Data segment of the nth slave station

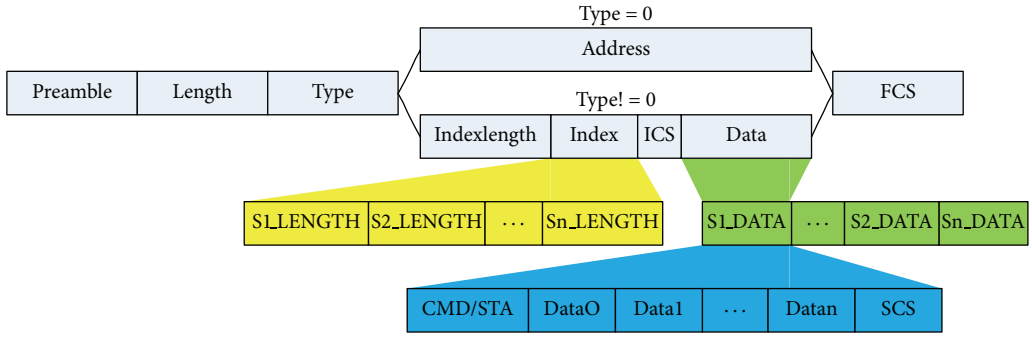


Figure 6: Frame format of motion control bus.

address allocation mechanism, rule is provided as follows: when address reset frames and data transmission frames are sent, function on retransmission and retransmission in advance can be used, while function on retransmission and retransmission in advance cannot be used in the process of allocating address.

5.2. Design and Realization of FPGA for Master Station and Slave Station. According to the protocol defined in this paper, function of FPGA from master station is relatively simple, and block diagram of the master station is shown in Figure 7.

Because the slave station needs to achieve straight-through forwarding function and generation function of synchronization signal, the slave function is relatively complex; block diagram of slave station is shown in Figure 8. The pass-through forwarding principle simultaneously receives and transmits, so receiving module of slave station and sending module of slave station are tightly coupled. Both of them link each other through sending strobe module shown in Figure 8.

5.3. The Overall Operation of the System and Reliability Verification. In order to verify the correctness of this protocol in this paper and design of FPGA of master and slave stations, this paper achieves protocol from Section 5.1 in system hardware platform. The configuration of hardware and software is set as follows.

(1) The master station and slave station use embedded motion controller body; software is based on operating system of μC/OS2.

(2) The slave station uses slave station verification board.

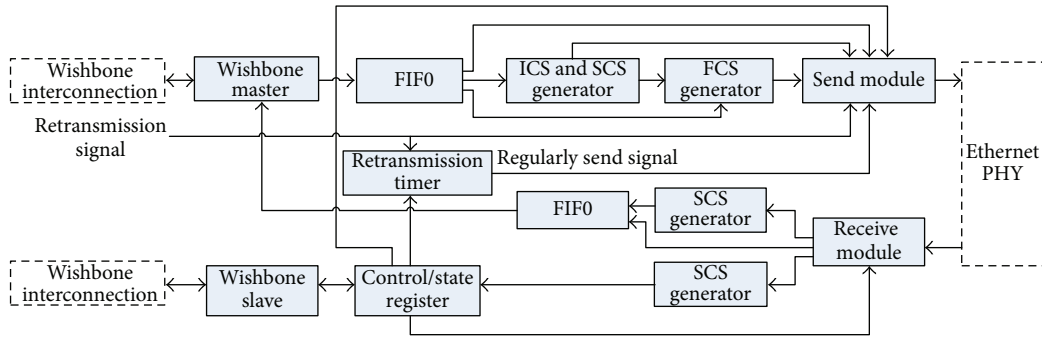


Figure 7: Block diagram of the master station.

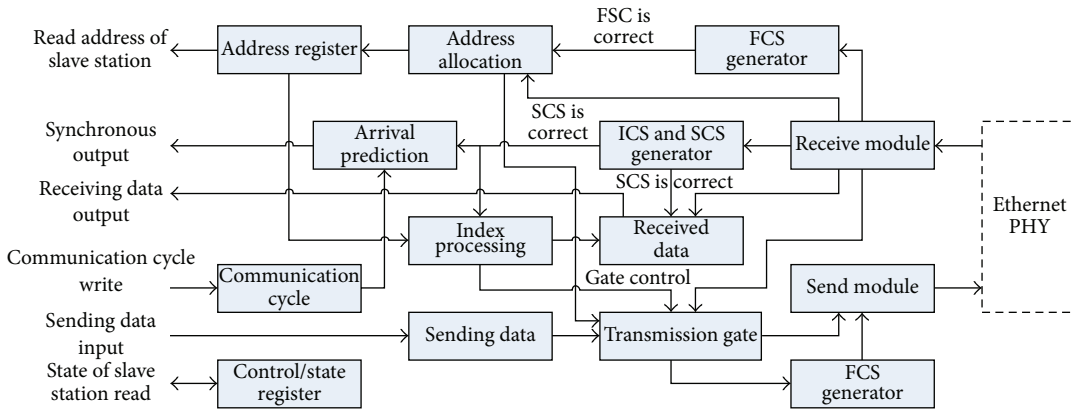


Figure 8: Block diagram of the slave station.

(3) Test system has 8 slave stations; when we use a single ring structure, the number of retransmission is 1. In order to verify the mechanism proposed in this paper in different working environments, the amount of data from slave station is used as standard; test conditions are divided into two groups, which are shown in Table 2.

Figure 9(a) is a debugging and synchronization performance testing, field environment of reliability testing. From Figure 9(b), synchronous jitter of the eight slave station is within 160 ns. Local clock of testing platform is 50 MHz, so results are in line with theoretical and simulation analysis from Section 2.2 in this paper.

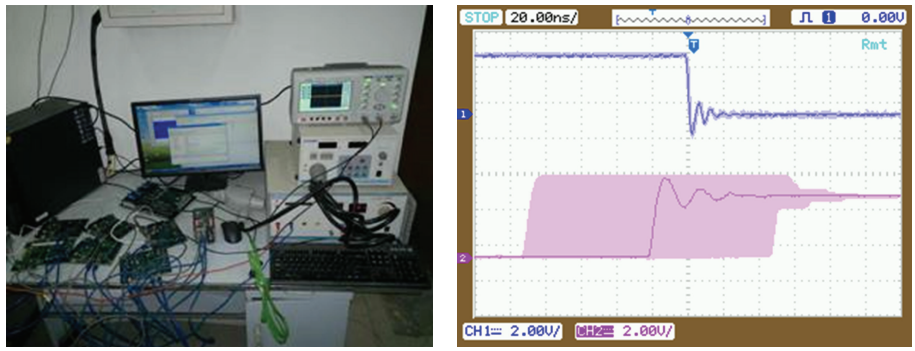
In order to verify the performance of independent verification mechanism of the data segment, this paper uses a burst interference device SKS-0404 and static equipment SKS-0230 to simulate power supply and electromagnetic interference in the actual industrial environment so as to increase the interference effect.

The reliability data is uploaded to PC for analysis by the master station through serials; successful communication criteria are defined as follows: (1) index segment verification ICS is correct; (2) Among the starting frame and the retransmission frame, one of the SCS came form data segment of all the slave stations is correct at random, and datum of the CMD/STA domain came from data segment are in line with the pre-specified format at the same time.

Table 2: Test system for the bus parameters.

Group for the amount of data	A	B
Effective datum of slave station (byte)	8	32
Frame length (byte)	124	316
Complete communication time of node frame (μs)	17.92	33.28
Minimum interval of frame retransmission (μs)	11.52	26.88
Actual interval of frame retransmission (μs)	13	30
Minimum communication cycle (μs)	30.92	63.28
Actual communication cycle (μs)	33.33	66.66

The test time was one hour, and test results are shown in Table 3. When we do not use independent verification of the data segment, two frames FCS are wrong, so it is considered that the communication is a failure, and the number in Group A test is 58; if the independent verification of data segment is used, the number of actual communication failure is 18 times; reliability has significantly improved. From the analysis of data communication failure, we can find that, in group A test ICS error, that is, parity error of index segments reaches 15,



(a) (b)

Figure 9: Synchronization performance and reliability testing site and synchronization waveform.

Table 3: Testing results for reliability of motion control bus.

Group for the amount of data	A	B
Communication times	108000000	54000000
FSC error by the first frame	112	178
FSC error by the retransmission frame	124	180
Error of two frame FCS	58	93
Two frame are all wrong (0/000)	0.0054	0.0172
Communication failure	18	12
communication failure (0/000)	0.0012	0.0022
The proportion of communication failure and wrong for two frames FSC	22.2%	12.9%
ICS error	15	7
The proportion of communication failure and ICS error	83.3%	58.3%

and index segment of independent verification mechanism of the data segment is communication failure factors, this situation is also consistent with the theoretical analysis. To work out this weak link, in some special networks, this paper proposes to use a protocol whose data segment has fixed length, and there is no need to use index segment in this circumstance, so salve station can directly find data segment by their own address, that is to say, reliability can be further improved.

From the comparison of the results of Group A and Group B, when the test time is the same, with the increase in the amount of the node data, the error probability is also increased, but increasing speed for error will be different if the mechanism proposed in this paper is used. The chance of error increases from the extreme of 0.0054 to 0.0172 with an increase of 319% when we do not use this mechanism. But by using the mechanism in this paper, the chance of error increases from the extreme of 0.0012 to 0.0022 with an increase of 183%, which illustrates that improvement effect for large datum transmission will be more obvious in this paper.

Analysis from above-mentioned ICS error shows that the ICS error is the main cause of communication failure; when the amount of data increases, the ICS error ratio decreases from 83.3% to 58.3%, which indicates that the mechanism in this paper has better performance when large amount of datum is transmitted.

6. Conclusions

In summary, this paper proposes the prediction mechanism in the arrival of the frames, the retransmission mechanism in advance, and a new independent verification of data segment. Tests show that the mechanism in this paper has better performance when large amount of datum is transmitted, and it is suitable for control system to transmit data.

Acknowledgment

This work is funded by the major projects of the National Science and Technology 04 special project: algorithm in whole digital and high-grade numerical control device and research on numerical control system field bus.

References

[1] P. Neumann, "Communication in industrial automation-what is going on?" *Control Engineering Practice*, vol. 15, no. 11, pp. 1332–1347, 2007.

[2] K. C. Lee and S. Lee, "Performance evaluation of switched Ethernet for real-time industrial communications," *Computer Standards and Interfaces*, vol. 24, no. 5, pp. 411–423, 2002.

[3] Z. Wang, Y. Q. Song, J. M. Chen, and Y. X. Sun, "Real time characteristics of Ethernet and its improvement," in *Proceedings of the 4th World Congress on Intelligent Control and Automation*, vol. 2, pp. 1311–1318, China, June 2002.

[4] J. D. Decotignie, "Ethernet-based real-time and industrial communications," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1102–1117, 2005.

[5] J. R. Moyne and D. M. Tilbury, "The emergence of industrial control networks for manufacturing control, diagnostics, and safety data," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 29–47, 2007.

[6] M. Labashov, G. Pratl, and T. Sauter, "Applicability of Internet pro-tools for field bus access," in *Proceedings of the 4th IEEE International Workshop on Factory Communication Systems*, pp. 28–30, 2002.

[7] M. Felser, "Real-time ethernet—industry prospective," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1118–1129, 2005.

[8] M. Felser, "Fieldbus based isochronous automation application," in *Proceedings of the IEEE Conference on Emerging Technologies and Factory Automation (ETFA '09)*, Mallorca, Spain, September 2009.

[9] J.-P. Thomesse, "Fieldbus technology in industrial automation," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1073–1101, 2005.

[10] J. D. Decotignie, "Ethernet-based real-time and industrial communications," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1102–1117, 2005.

[11] H. F. Abdel-Ghaffar, M. F. Abdel-Magied, M. Fikri, and M. I. Kamel, "Performance analysis of fieldbus in process control systems," in *Proceedings of the American Control Conference*, pp. 591–596, June 2003.

[12] IEC, "Real-Time Ethernet: EPL, (Ethernet Powerlink): proposal for a publicly available specification for real-time ethernet," Tech. Rep., Doc. IEC65C/356a/NP, 2004.

[13] G. Cena, L. Seno, A. Valenzano, and S. Vitturi, "Performance analysis of Ethernet Powerlink networks for distributed control and automation systems," *Computer Standards and Interfaces*, vol. 31, no. 3, pp. 566–572, 2009.

[14] J. Eidson and K. Lee, "IEEE 1588 Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," in *Proceedings of the 2nd ISA/IEEE Sensors for Industry Conference*, pp. 98–105, Houston, Tex, USA, 2002.

[15] J. Jasperneite, M. Schumacher, and K. Weber, "Limits of increasing the performance of industrial ethernet protocols," in *Proceedings of the 12th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '07)*, pp. 17–24, Patras, Greece, September 2007.

[16] G. Prytz, "A performance analysis of EtherCAT and PROFINET IRT," in *Proceedings of the 13th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA '08)*, pp. 408–415, Hamburg, Germany, September 2008.

[17] M. Schumacher, J. Jasperneite, and K. Weber, "A new approach for increasing the performance of the industrial ethernet system PROFINET," in *Proceedings of the 7th IEEE International Workshop on Factory Communication Systems (WFCS '08)*, pp. 159–167, Dresden Germany, May 2008.

[18] G. Cena, S. Scanzio, A. Valenzano, and C. Zunino, "Performance evaluation of the EtherCAT distributed clock algorithm," in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE '10)*, pp. 3398–3403, Bari, Italy, July 2010.

[19] D. Gunzinger, C. Kuenzle, A. Schwarz, H. D. Doran, and K. Weber, "Optimising PROFINET IRT for fast cycle times: a proof of concept," in *Proceedings of the IEEE International Workshop on Factory Communication Systems (WFCS'10)*, pp. 35–42, Nancy, France, May 2010.

[20] M. Knezic, B. Dokic, and Z. Ivanovic, "Topology aspects in EtherCAT networks," in *Proceedings of the 14th International Power Electronics and Motion Control Conference (EPE-PEMC '10)*, pp. T11–T16, Ohrid, Macedonia, September 2010.

[21] M. Knezic, B. Dokic, and Z. Ivanovic, "Increasing EtherCAT performance using frame size optimization algorithm," in *Proceedings of the 16th IEEE Conference on Emerging Technologies & Factory Automation (ETFA '11)*, pp. 1–4, Toulouse, France, 2011.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 824719, 8 pages
http://dx.doi.org/10.1155/2013/824719

Research Article

A Force-Position Control System Based on Soft Tissue under Large Deformation

Lei Hu,[1,2] Jie Zhang,[2] Changsheng Li,[2] Yan Cui,[2] and Liquan Wang[1]

[1] *School of Mechanical and Electrical Engineering, Harbin Engineering University, Harbin 150001, China*
[2] *School of Mechanical Engineering and Automation, Beihang University, XueYuan Road No. 37, HaiDian District, Beijing 100191, China*

Correspondence should be addressed to Changsheng Li; lics32@163.com

Received 21 April 2013; Accepted 20 May 2013

Academic Editor: Hongxing Wei

We introduce an extension of the mass-spring model applied to human tissue that can truly reflect the mechanical properties of a living person's tissue under large deformations under pressure by a massage head. The model is used in the force-position control system and helps us to control the massage force more accurately. Parameters of the model were determined according to the experimental results. Simulation results demonstrated that the model succeeded in modeling living tissue under large deformation, and the force-position control system based on the soft tissue model under large deformation achieved higher accuracy than the control system based on the soft tissue model with constant stiffness.

1. Introduction

Chinese massage has been found to be an efficient way to ease and treat chronic pain [1]. With modernization, massage robots have been used with success in place of traditional massage techniques. Japan's Sanyo Electric Co. first utilized robotic kneading action in 1996 as an alternative for hand massage [2]. Toyohashi University of Technology proposed a four-finger massage hand with force sensor in 2006 [3–6]. Tottori University subsequently introduced a massage chair based on skin resilience in 2007 [7], and in the same year, Japan's Waseda University and Asahi University developed robots for facial massage [8], and Konkuk University applied the robot to back massage [9]. Although these studies discussed the specific aspects of each massage, the entire massage system that integrates massage skills together with the therapy process needs further study. In addition, safety considerations and individualized massage therapy solutions should be incorporated into robotic massage system, so as to make them more adaptable for clinical use.

For these reasons, we developed a robotic massage system acting on acupuncture points based on traditional Chinese massage therapy theory, with reproduced human manipulation technics based on the experience of experts [10]. The kinematic and force features of key massage technics, such as thumb kneading, pressing, rolling, vibrating, and pinching, were summarized by analyzing the massage processes of massage therapists, and a mathematical model for robotic massage was established. A pain threshold value was introduced to individualize therapy schemes, and a force-position control method based on the pain threshold was presented. VAS (Visual Analogue Scale) tests for lumbar muscle strain were carried out using the massage robot to validate the treatment effect.

The robot end-effector was contacted with subjects' soft tissue when working, and the control system was affected, which allowed the proposal of a system based on soft tissue [11]. However, the system was not stable across individuals because the soft tissue model was simplified as a spring-damping system with constant stiffness, which is only applied to people under small deformations. To achieve more precise control of massage strength, soft tissue should be modeled under large deformations. Therefore, this paper presents a method to model living tissue under large deformations and a corresponding control system.

2. System Overview

The system is composed of a massage bed, a positioning platform, a massage end-effector, a visual tracking system,

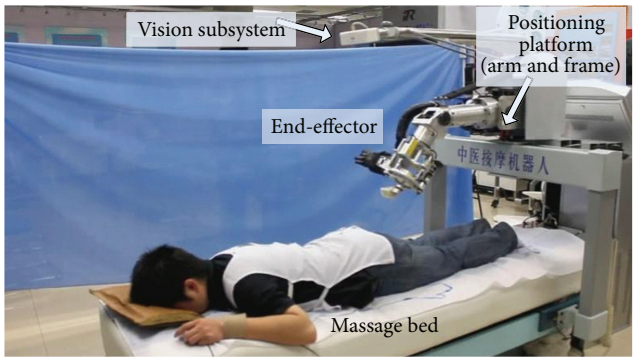


Figure 1: System configuration of the massage robot.

and auxiliary devices, as shown in Figure 1. The massage bed is the structural base of the whole system, and the patient lies on it during robotic massage. Auxiliary devices, such as audio players, physiological information-collecting devices, and pain threshold acquisition devices, are embedded in or attached to the bed [10].

The positioning platform is mounted on the massage bed and is designed to position the end-effector on the massage spot and fulfill manipulation technics when needed. As the positioning platform runs over the patient's back, it consists of two assemblies, that is, a gantry like a moving frame for translational positioning in three perpendicular directions and an arm with four rotational joints so that the position and orientation of the end-effector could be adjusted for different techniques at specific acupuncture points. The maximum workspace of the platform is $180 \times 55 \times 60$ cm, which allows the end-effector to reach every part of patient's body to perform the massage. However, it is difficult to integrate all the different manipulation techniques for various therapy functions, such as rolling, pressing, vibrating, thumb kneading, and pinching into one mechanical manipulator. Therefore, the end-effector is designed to be composed of two parts, that is, a parallel mechanism with five degrees of freedom (DOFs) to realize rolling, pressing, and vibrating techniques and a humanoid massage hand with one DOF for thumb kneading and pinching techniques. According to the specific requirements for different techniques, the massage techniques are reproduced by the end-effector. The visual-tracking device of a stereo camera is installed on an elongated arm extending from the positioning platform and moves along with the platform. Markers are attached to both the patient body and the robot body, so the tracking device can calculate the target position relative to the robot coordinate frame. With its scope of view covering the workspace of the end-effector, the tracking device performs real-time monitoring in case the patient moves; thus, it can ensure safety and precision of the massage operation. The tracking error of the massage robot's vision servosystem is less than 2.5 mm.

An audio device is embedded into the massage bed to help the patient relax, which is advantageous for augmenting the treatment effect according to traditional Chinese massage therapy theory. Physiological information-collecting devices are also provided as an input to the robotic control system for adaptive control according to the patient's current physiological status. The pain threshold acquisition device is a digital glove equipped with force sensors. Before beginning massage, physicians use this glove to press pain points on the patient to determine their pain threshold value, which is sent to the control system to customize the massage force for the specific patient.

The main control strategy of the system is force-position control; it has two control modes in the massage operation process. The position-tracking control is used in free space before the end-effector contacts with the human body. The other is force-tracking control used in restricted space after the end-effector contacts the human body. The two control modes can be shifted according to the contacting force magnitude. Section 4.2.2 describes the errors of force-position control, which are less than 0.05 N and 0.1 mm.

Moreover, safety elements were designed under the consideration that the robot would be in close contact with the patient. For instance, motion DOF redundancy is provided in the positioning platform to avoid interference of the robot body with patients. There are seven DOFs in the positioning platform, and motion sequences are configured to proceed from one axis to another; the vertical prismatic joint is driven after all the other joints move.

3. Material and Methods

3.1. Modeling Living Tissue under Large Deformation Using an Index Model

3.1.1. Experimental Setup. To establish an accurate human soft tissue model, we needed to perform experiments on the soft tissues of living subjects. Properties of the adipose tissue (mainly fat) have been characterized using compression tests of human breast [12–15], heel [16, 17], and forearm, [18], as well as pig tissue [19, 20]. In this paper, we designed an experimental setup to measure the displacement of living tissue when different forces compress a person's arm. Because our experiment was associated with the Chinese massage robot project, study subject safety was assured. There are very few data available in the literature regarding the properties of living human tissue. Furthermore, precise steps were taken to adjust the parameters of the simulation model for further clinical applications.

The experiment setup for measuring the mechanical properties of living tissue is shown in Figure 2. A 6 mm diameter tip was mounted on a force sensor. The velocity was controlled by a servomotor with a large range of speed. The external force on human tissue was obtained by the force sensor, and the three-dimensional coordinates of the marker on human's arm were later written to a file when the tracker communicated to a PC by A/D. The data of the force and the vertical movement of the tip were used to generate the force-displacement curve.

3.1.2. Experiment Results. To obtain the force-displacement curve, the tip was moved at a constant velocity to compress the person's arm until they felt pain. At the same time, the force applied to the person's arm and the tip displacement

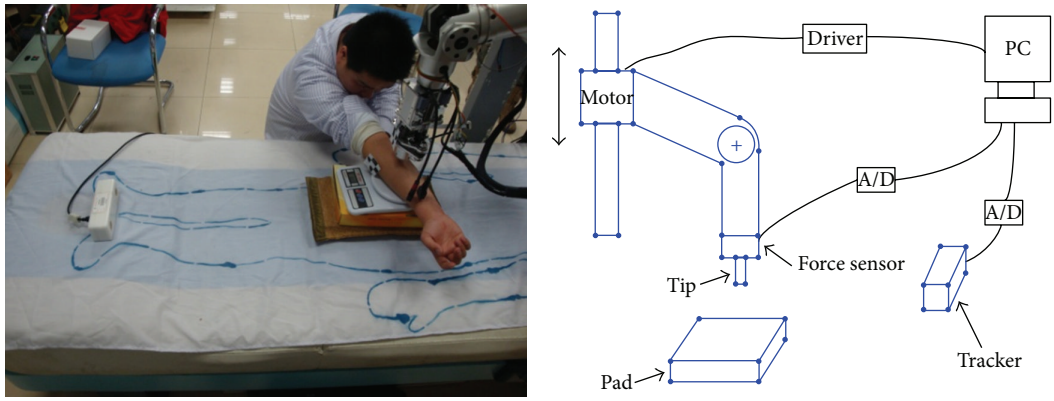


Figure 2: Experimental setup for measuring mechanical properties of living tissue.

were measured and recorded at a certain frequency. Five different velocities were considered: 5, 10, 15, 20, and 30 mm/s. Five independent environments were carried out for each velocity. Figure 3 shows the displacement-force relationship curves for one person at five different velocities. The figure illustrates that the displacement-force relationship curves were very similar, so we used the velocity of 30 mm/s to study the human soft tissue model.

3.1.3. Index Model Method

(1) Mass-Spring Model. Mass-spring system is based on physical law. It has been widely and efficiently used in modeling various deformed objects. In this system, the modeled object is discretized into cubes, each containing four tetrahedrons. Each point is in 4 cubes and has 18 points surrounding it [21].

Usually, the spring's elasticity is linear. In the dynamic system, the motion of each single quality key point follows Newton's second law [22], which is expressed by

$$m_i \ddot{x}_i = -\gamma_i \dot{x}_i + \sum_j g_{ij} + f_i, \tag{1}$$

where m_i is the mass of point i and x_i is a three-dimensional vector presenting the position of point i. $\dot{x}_i$ and $\ddot{x}_i$ represent the first and the second derivation of x_i to time t. On the right of the formula are the forces applied to point i. The first term is the resistance related to the speed, γ_i is the damping coefficient. $\sum_j g_{ij}$ is the sum of all forces contributed by all adjacent points of i. The third term is the total external force applied to point i.

The motion equation of the entire system is obtained by integrating all the motion equations of each point in the grid structure. By summing N position vectors of each independent quality point to a $3N$-dimensional vector, we obtain an equation that takes the form of

$$M\ddot{x} + C\dot{x} + Kx = f, \tag{2}$$

where M, C, K are all tensors of size $3N \times 3N$, representing the mass, resistance, and the elastic coefficient. x is a $3N$-dimensional position vector. Although these matrices are very large, they are also sparse. M and C are diagonal matrices, and K is a banded matrix because the spring tension

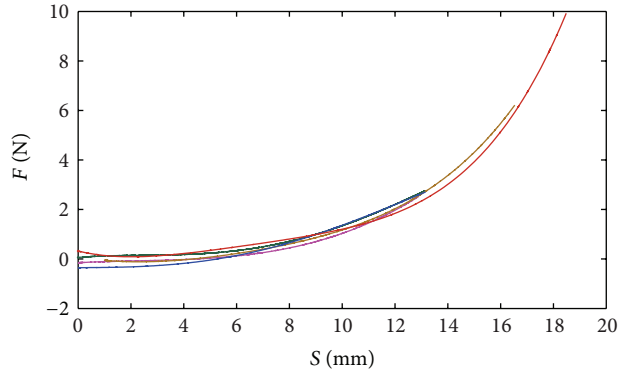


Figure 3: Displacement-force relationship curve.

depends only on the distance between the adjacent quality points. f is a $3N$-dimensional vector, representing the total external forces applied to all the quality points. Equation (2) can be transformed into two first-order differential equations expressed by

$$\dot{v} = M^{-1}\left(-Cv - Kx + f\right),$$
$$\dot{x} = v, \tag{3}$$

where v is the velocity vector of the entire quality point system. Then the x and v function expressions of time t can be solved by using different integral methods.

To save computation time, we chose an explicit scheme based on the Euler method to solve the dynamic system. Vertex position at time $t + h$ is derived from the position at time t using the following equations [21]:

$$x_i(t+h) = x_i(t) + h * v_i(t),$$
$$a_i(t) = \left(F_i^{\text{ext}}(t) + \sum_{j \in s_i} K_{i,j} * \frac{\left|x_j(t) - x_i(t)\right| - L_{i,j}}{L_{i,j}} * \frac{x_j(t) - x_i(t)}{\left|x_j(t) - x_i(t)\right|} - \gamma_i * v_i(t)\right) \times (m_i)^{-1},$$
$$v_i(t+h) = v_i(t) + h * a_i(t), \tag{4}$$

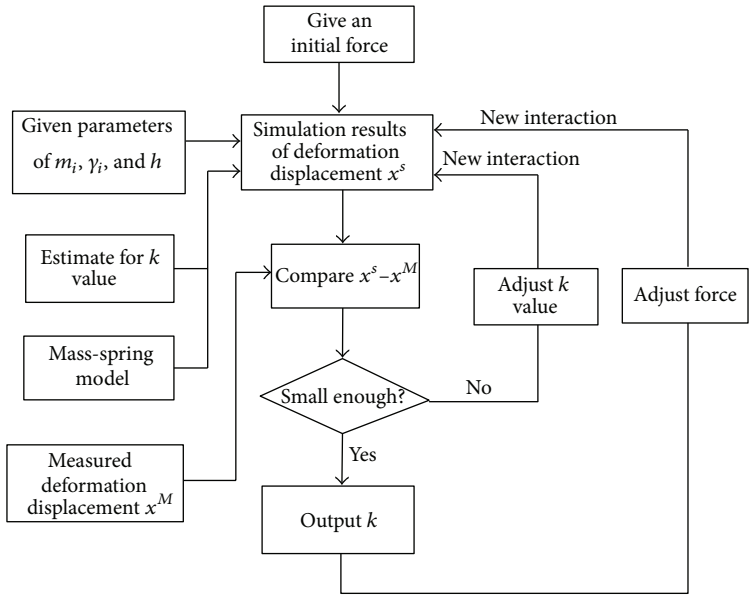


Figure 4: k value calculation.

where $a_i(t)$, $v_i(t)$, and $x_i(t)$ are the acceleration, velocity, and the position of point i; F_i is the external force of point i; $K_{i,j}$ is the elastic modulus of the spring connecting points i and j; $L_{i,j}$ is the free length of the spring; and γ_i and m_i are the damping coefficient and the mass of point I, respectively.

The convergence condition is when acceleration and velocity are the lowest. In practice, the system achieves balance when the acceleration and velocity are less than specific given values.

(2) Index Modeling Method. Previous models were based on the assumption of linear elasticity theory. However, the results of our experiment (Section 3.1.2) showed that the deformation of live persons' soft tissues was not linearly elastic. To establish the most suitable model for modeling living tissue under large deformations, the expression of the elastic coefficient in formula (2) must be determined. Because human tissue has viscoelasticity, the deformation should have an exponential delay ($e^{-t/\tau}$) after the force is exerted on or removed from human tissue [23]. We found that the force-displacement curve was similar to the index curve. Because the derivatives of the index function are also the index function, we supposed that the elastic coefficient k is the index function of displacement. The elastic coefficient k in (2) can be expressed as follows:

$$k = a \cdot e^{bx}, \tag{5}$$

where a and b are the parameters to be determined and x is the deformation displacement according to the experiment results. Then, just as in the mass-spring model, the system can be solved using the Euler method.

In order to find the k expression with regard to deformation displacements, we used the Euler method to calculate k value corresponding to different forces. As shown in Figure 4, the calculating steps for each force are as follows.

(i) An initial force was given.

(ii) The parameters of m_i, γ_i, and h were given, and the k value was estimated.

(iii) Take the parameters above into the formula and use the Euler method to simulate.

(iv) Compare the simulation and experimental results. If the difference is smaller than a certain value, then output k value or adjust k value and simulation from (ii) again until a suitable k value is obtained.

(v) Change the value of force and jump to (ii).

3.2. Control System Based on the Soft Tissue under Large Deformation. A control system based on the soft tissue model under large deformation mentioned in Section 3.1 is proposed. The soft tissue is simplified as a spring-damping system, which is expressed by

$$F(t) = Kx(t) + B\dot{x}(t), \tag{6}$$

where $F(t)$ is the contact force between the end-effector of the massage robot and the soft tissue, $x(t)$ is the position of the end-effector, K is the stiffness, and B is the damping coefficient.

Calculating the Laplace transform of (6) yields the following equation:

$$F(s) = (K + B)X(s). \tag{7}$$

The values of K and B can be determined depending on the simulation results in Section 3.1, and they can be expressed by

$$\begin{aligned} B &= 51 \times 51 \times \gamma, \\ K &= 51 \times 51 \times k. \end{aligned} \tag{8}$$

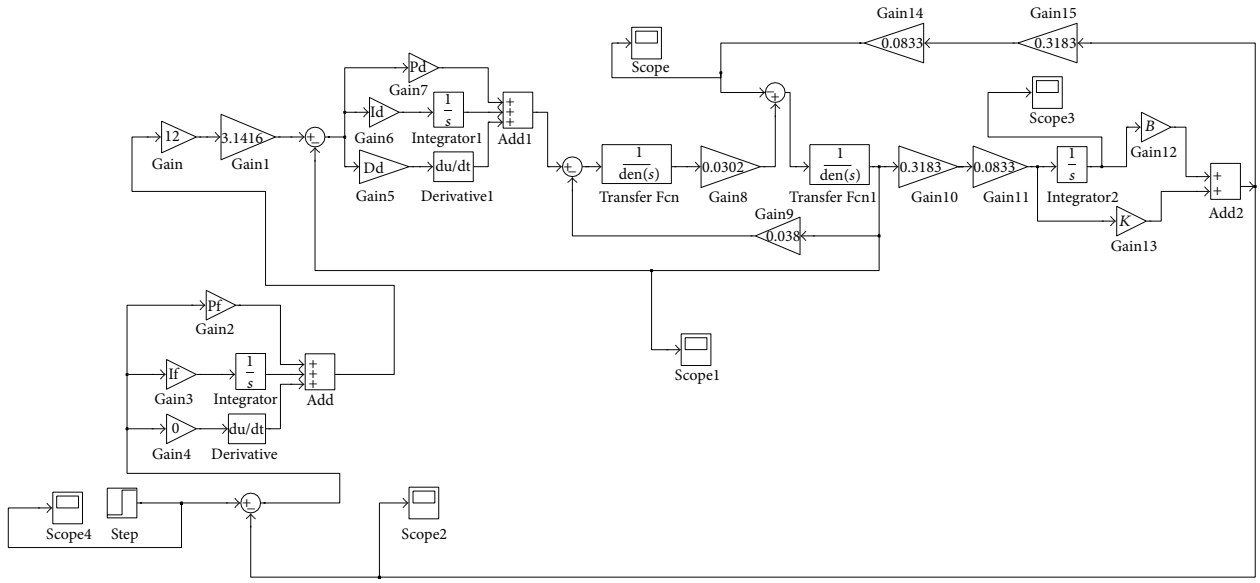


FIGURE 5: Z-axis motor control system.

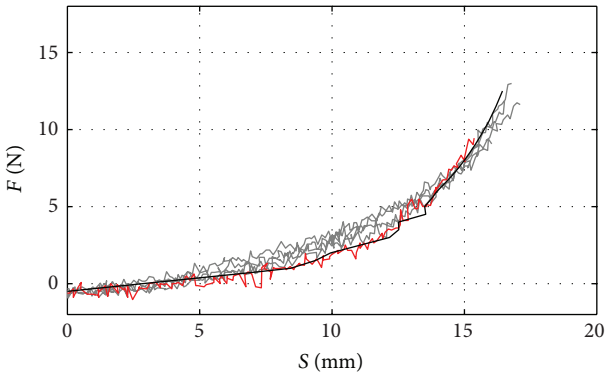


FIGURE 6: Comparison between simulation and experiment results. Simulation results are in black, the referred experiment results are in red, and the other four experiment measurements at the same speed of 30 mm/s are in gray.

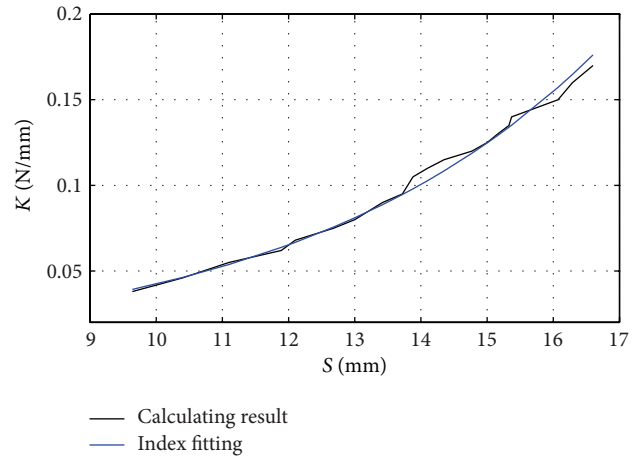


FIGURE 7: Comparison between fitting and calculating results. Calculating results of k value are in black, and index function fitting results are in blue.

The relationship between the motor output torque and the voltage can be expressed by

$$T_m(s) = \frac{U_a(s) - c_a w_m(s)}{L_a(s) + R_a} c_m(s). \quad (9)$$

In the system, the force controller is controlled by Pf and If, while the displacement controller is controlled by Pd, Id, and Dd.

The Z-axis motor control system was modeled using the Simulink toolbox in Matlab, as expressed in Figure 5.

4. Results and Discussion

4.1. Index Modeling Results. Only some types of human tissue are deformed when force is applied to the human arm. So 51× 51 × 40 vertices and 97,500 cubes were used in the system for the test. The sides of all these cubes are 1 mm. To ensure that the simulation results are convergent, after a few tests, the parameters were given as follows: $m_i = 0.001$ g (according to human density), $\gamma_i = 0.0005 \cdot 10^{-3}$ N·S/m, $h = 0.01$ s, and the two convergence thresholds were both 0.0005.

Figure 6 shows the comparison between simulation and experimental results. The black curve in Figure 4 plots the simulated deformation displacement under different forces obtained after parameter fitting, the red curve is the experiment measurement used to calculate the elastic coefficient, and the gray curves are the experiment measurements of four different people at the same speed of 30 mm/s. The simulation shows good correspondence to the red experimental curve, and the trend of the experiment curves similar to all four subjects.

Table 1: Simulation comparison results.

F/N	1	2	3	4	5	6	7	8	9
$S1$/mm	8.6896	9.9223	11.4006	12.4641	13.5718	14.2399	14.6321	15.0191	15.3401
$S2$/mm	3.6144	6.2376	8.3779	9.8444	12.1991	14.3311	15.2253	15.9305	16.4356
$S3$/mm	8.5	10.4	11.89	12.68	13.42	13.88	14.35	15	15.33

Table 2: Error comparison results.

F/N	1	2	3	4	5	6	7	8	9
Error 1/mm	0.1896	0.4777	0.4894	0.2159	0.1518	0.3599	0.2821	0.0191	0.0101
Error 2/mm	4.8856	4.1624	3.5121	2.8356	1.2209	0.4511	0.8753	0.9305	1.1056

Table 3: K values under different forces.

F/N	1	2	3	4	5	6	7	8	9
K/N/mm	79.85	120.16	165.94	196.64	230.7	254.9	281.9	324.3	348.3

Table 4: Force data accuracy.

$F1$/N	1	2	3	4	5	6	7	8	9
$F2$/N	1.05	1.98	3.02	4.01	5.0	5.96	7.05	8.01	9.03
Error/N	0.05	0.02	0.02	0.01	0	0.04	0.05	0.01	0.03
Accuracy	5%	1%	0.67%	0.25%	0%	0.67%	0.7%	0.14%	0.33%

Figure 7 shows the comparison between the fitting and calculating results of the elastic coefficient. The black curve in Figure 5 shows the calculating results of k value corresponding to deformation displacements, while the blue curve shows the fitting curves using index function. The data show that the curve using index function fits closely to the calculating data.

The relationship between force and displacement using the constant stiffness model when $k = 0.1$ and the index model mentioned above were calculated. The simulation comparison results are shown in Table 1; $S1$ and $S2$ are the simulation results using the index model and the constant stiffness model, and the $S3$ row shows the measurement results. The error comparison results are shown in Table 2, of which Error 1 and Error 2 are the errors between the measurement results and the simulation results using the index model and the constant stiffness model, respectively.

From the above tables, we can see that the maximum error between simulation results and measurement results can be decreased from 4.8856 to 0.4894, and the accuracy can be improved from 57.5% to 4.6%. These findings show that the index model succeeded in more accurately expressing the mechanical property of living tissue under large deformation.

4.2. Control System Results

4.2.1. Force Simulation Results. By imputing the simulation results into (8), we can get $B = 1.3$ NS/mm, and the K values under different forces are shown in Table 3.

Figure 8 shows the force simulation results. From the figure, we can see that the both of the errors are less than 0.1% and the system has good stability.

4.2.2. Force Measurement Results. Using the experimental setup mentioned in Section 3.1.1, different force values were input to the massage robot, and we compared the robot's given forces ($F1$) and actual forces ($F2$). The results are shown in Table 4.

From Table 4, we can see that the actual force is quite close to the given force. The error was less than 5%, which is more accurate than the previously reported value of 10% [11].

Above all, the force-position control system based on soft tissue under large deformation succeeded in accurately controlling the massage force imposed on people.

5. Conclusion

We designed an experiment setup to measure the displacement of living tissue when different forces compressed a person's arm. We introduced an index model that can accurately reflect the biomechanical characteristics of living tissue under large deformation.

The model allowed us to improve the force-position control system using the simulation results of K and B. The experimental results demonstrate that the control system has succeeded in following the tracks of the given force, and the error is less than the control system based on soft tissue

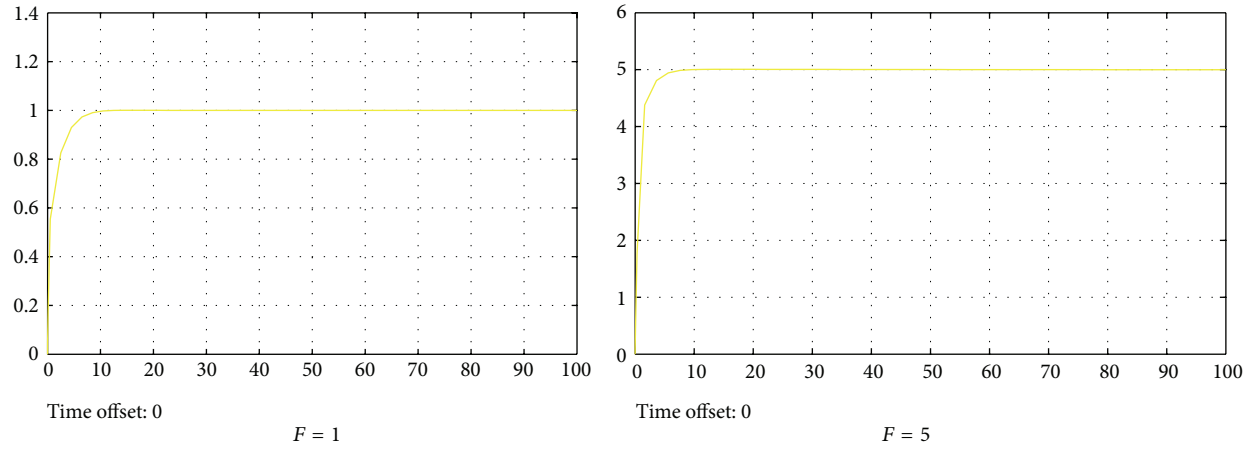


Figure 8: Force simulation results.

with constant stiffness. Moreover, because the K and B were obtained from the proposed soft tissue model when different forces were imposed, the control system is more stable for making estimates in different people.

Acknowledgment

This project was supported by the National Support Project Chinese Massage Robot.

References

[1] S. Alavandar, K. A. V. Sundaram, and M. J. Nigam, "Genetic algorithm based robot massage," *Journal of Theoretical and Applied Information Technology*, vol. 3, no. 4, pp. 102–109, 2007.

[2] T. Field, M. Diego, and M. Hernandez-Reif, "Massage therapy research," *Developmental Review*, vol. 27, no. 1, pp. 75–89, 2007.

[3] V. Golivin and A. Grib, "Mechatronic system for manual therapy and massage," in *Proceedings of the 8th Mechatronics Forum Intemational Conference (Mechatronics '02)*, pp. 774–784, University of Twente, June 2002.

[4] C. Innocenti and P. Wenger, "Position analysis of the RRP-3(SS) multi-loop spatial structure," *Journal of Mechanical Design, Transactions of the ASME*, vol. 128, no. 1, pp. 272–278, 2006.

[5] P. Minyong, T. Miyoshi, K. Terashima, and H. Kitagawa, "Position and force control of multi-fingered robot hand and its application to massage motion," in *Proceedings of the IASTED International Conference on Robotics and Applications*, pp. 100–105, June 2003.

[6] K. Mouri, K. Terashima, P. Minyong, H. Kitagawa, and T. Miyoshi, "Identification and hybrid impedance control of human skin muscle by multi-fingered robot hand," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '07)*, vol. 1–9, pp. 2895–2900, San Diego, CA , USA, November 2007.

[7] M. Lewis and M. I. Johnson, "The clinical effectiveness of therapeutic massage for musculoskeletal pain: a systematic review," *Physiotherapy*, vol. 92, no. 3, pp. 146–158, 2006.

[8] R. C. Luo, C. C. Chang, and Y.-W. Perng, "Impedance control on a multi-fingered robot hand based on analyzed electromyographic information for massage applications," in *Proceedings of the IEEE International Symposium on Industrial Electronics (IEEE ISIE '09)*, pp. 1217–1222, Seoul, Korea, July 2009.

[9] K. Terashima, H. Kitagawa, T. Miyoshi, P. Minyong, and T. Kondo, "Modeling and massage control of human skin muscle by using multi-fingered robot hand," *Integrated Computer-Aided Engineering*, vol. 13, no. 3, pp. 233–248, 2006.

[10] H. lei, W. Yun, Z. Jie et al., "A Massage robot based on Chinese massage therapy," *Industrial Robot*, vol. 40, no. 2, pp. 158–172, 2012.

[11] Y. Xu and T. Yao, "Simulation and implement of traditional Chinese finger-kneading based on hybrid force position control," in *Proceedings of the 2nd International Conference on Information Science and Engineering (ICISE '10)*, pp. 5384–5387, Hangzhou, China, December 2010.

[12] J. H. Chung, V. Rajagopal, P. M. F. Nielsen, and M. P. Nash, "A biomechanical model of mammographic compressions," *Biomechanics and Modeling in Mechanobiology*, vol. 7, no. 1, pp. 43–52, 2008.

[13] A. Samani, J. Bishop, M. J. Yaffe, and D. B. Plewes, "Biomechanical 3-D finite element modeling of the human breast using MRI data," *IEEE Transactions on Medical Imaging*, vol. 20, no. 4, pp. 271–279, 2001.

[14] A. Samani and D. Plewes, "A method to measure the hyperelastic parameters of *ex vivo* breast tissue samples," *Physics in Medicine and Biology*, vol. 49, no. 18, pp. 4395–4405, 2004.

[15] L. Z. Sun, Z. G. Wang, Y. Liu, and G. Wang, "Elastography method for reconstruction of nonlinear breast tissue properties," *International Journal of Biomedical Imaging*, vol. 2009, Article ID 406854, 9 pages, 2009.

[16] C. Flynn and B. A. O. Mccormack, "Finite element modelling of forearm skin wrinkling," *Skin Research and Technology*, vol. 14, no. 3, pp. 261–269, 2008.

[17] J. E. Miller-Young, N. A. Duncan, and G. Baroud, "Material properties of the human calcaneal fat pad in compression: experiment and theory," *Journal of Biomechanics*, vol. 35, no. 12, pp. 1523–1531, 2002.

[18] J. T. Iivarinen, R. K. Korhonen, P. Julkunen, and J. S. Jurvelin, "Experimental and computational analysis of soft tissue stiffness in forearm using a manual indentation device," *Medical Engineering and Physics*, vol. 33, no. 10, pp. 1245–1253, 2011.

[19] Y.-P. Zheng and A. F. T. Mak, "An ultrasound indentation system for biomechanical properties assessment of soft tissues *in-vivo*,"

IEEE Transactions on Biomedical Engineering, vol. 43, no. 9, pp. 912–918, 1996.

[20] J. Z. Wu, R. G. Cutlip, M. E. Andrew, and R. G. Dong, "Simultaneous determination of the nonlinear-elastic properties of skin and subcutaneous tissue in unconfined compression tests," *Skin Research and Technology*, vol. 13, no. 1, pp. 34–42, 2007.

[21] C. Jiang, *Research on Elastic Deformable Model of Soft Tissue*, Southeast University, 2004.

[22] S. F. F. Gibson and B. Mirtich, "A survey of deformable modeling in computer graphics," MERL, TR-97-19, 1997, http://www.merl.com.

[23] W. LIjuan, *The Human-Computer Interaction Control System Study of Force Feedback in the Virtual Surgery*, Shanghai Jiaotong University, Shanghai, China, 2010.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 974152, 9 pages
http://dx.doi.org/10.1155/2013/974152

Research Article

Look-Ahead Algorithm with Whole S-Curve Acceleration and Deceleration

Youdong Chen, Xudong Ji, Yong Tao, and Hongxing Wei

School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China

Correspondence should be addressed to Youdong Chen; chenyd@buaa.edu.cn

Received 5 April 2013; Accepted 12 May 2013

Academic Editor: Shao Zili

Tool paths of a complex contour machining generated by commercial CAD/CAM systems are mainly composed of many short linear/circular blocks. Though the look-ahead algorithms can improve speed and accuracy in the machining of short linear/circular segments, most of them just deal with linear segments with trapezoid acceleration and deceleration (acc/dec). In addition, the look-ahead algorithms with S-curve acc/dec are too complex to adopt the equivalent S-curve profile by approximation algorithm. To increase the smoothness of feedrate profile and machining efficiency of continuous short line and circle machining, this paper presents a feedrate profile generation approach and corresponding look-ahead algorithm with whole S-curve acc/dec. With the proposed look-ahead scheme, the feedrate profiles with S-curve acc/dec can work efficiently in a lot of short line and circle segments. Thus, the machining productivity can be increased and the feedrate profiles are smooth. The simulation and experiments verify the feasibility and validity of the proposed approach.

1. Introduction

Tool paths are currently generated by commercial CAD/CAM systems. Curves are usually approximated to a number of linear/circular blocks in CAD/CAM systems according to the feedrate and accuracy in CNC machining. During high-speed machining, the rapid and abrupt change of direction at the corner between two adjacent blocks causes vibration which is harmful to the machine mechanics and the surface finish. Therefore, to maintain the machining quality, the feedrate at the corner is set to zero in conventional NC machining, which may lead to low efficiency.

To maintain the machining quality and improve the efficiency with multiple paths within machining parameter constraints, many approaches have been proposed. One of approaches is parametric interpolation algorithms [1, 2], which generates a curved cutter path directly without segmentation contour processing. However, the computation load is heavy and inevitable errors are introduced by employing a truncated Taylor series. At the same time, there are a lot of actual workpieces described by the linear and circular codes. Therefore, there is a need to research the linear and cicular interpolation algorithm.

One of the most efficient approaches is look-ahead interpolators. In order to keep the feedrate continuity, it looks ahead several paths to detect the sharp corners of paths and overrides the feedrate profile instead of planning the individual path with an acc/dec profile. References [3–7] proposed an approximation method which fit continuous short line blocks into parametric curves before interpolation. Feedrate fluctuations associated with the approximation errors should be limited in this method. Zhang et al. [8] presented a multiperiod turning method to improve the feedrate at the junctions using the linear acceleration and deceleration mode. But this method is only applied to micro line segment. Some velocity planning approaches have been proposed in [9–12]. All these methods used linear acc/dec algorithm, which usually exhibits a jump in acceleration at the beginning and the end of velocity adaptation phase. S-curve acc/dec with better servo instantaneous characteristic is used in look-ahead algorithm [13, 14]. However, it is too complex to adopt the equivalent S-curve profile by approximation algorithm. It is not a whole S-curve profile, which may lead to overshoot for large acceleration.

It should be noted that the path lengths are arbitrary in the usual NC programming. All these look-ahead algorithms

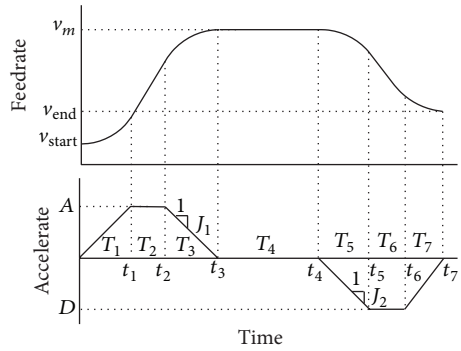


Figure 1: S-curve feedrate profile.

mentioned are only used for continuously microline segments. This paper proposed a look-ahead interpolation algorithm with the S-curve acc/dec for continuous linear/circular segments. During the motion, the feedrate profiles are constructed and checked within the machining constraints and the discretization of path with S-curve acc/dec. In addition, the paths of arbitrary length are also considered during interpolation.

This paper is organized as follows. In the next section, feedrate profile formulations with S-curve acc/dec are derived. Section 3 details the look-ahead algorithm. Simulation and experiments with the proposed look-ahead scheme are given in Section 4. Finally, a conclusion is given in Section 5.

2. Feedrate Profile Formulation

2.1. Discretizing the Normal S-Curve acc/dec. S-curve acc/dec can avoid mechanical shock imposed on the servo system and get stable motion along the tool path. Figure 1 shows the kinematics profiles used in the normal S-curve acc/dec. The profiles can be divided into acceleration-speed zone, constant-speed zone, and deceleration-speed zone. For the motion along the tool path, accelerations have trapezoidal profiles with prespecified slopes values. The maximum allowable acceleration and deceleration are A and D, respectively.

Considering Figure 1, the starting point feedrate is v_{start}, the end point feedrate is v_{end}, the desired feedrate is v_m, the starting and end point acceleration is zero, the maximum allowable jerk at acceleration-speed zone is J_1, and the maximum allowable jerk at deceleration-speed zone is J_2. The normal S-curve acc/dec has seven regions with time T_i ($i = 1, 2, \ldots, 7$), $T_1 = T_3 = n_a T_s$, $T_5 = T_7 = n_b T_s$; $T_4 = n_c T_s$, $T_1/T_2 = T_5/T_6 = k$, n_a, n_b, n_c and k are integral, and T_s is the sample period. The travel length is L. If the feedrate profile is symmetrically smooth, n_a is equal to n_b. In the traditional CNC system, the times T_1, T_5 and k are prespecified. If there is acceleration or deceleration, the acceleration time or deceleration time are the prespecified constants, which may reduce efficiency. The accelerations A and D and the jerks J_1 and J_2 are prespecified in terms of machine mechanics.

2.1.1. Determination of the Numbers of Interpolation Steps for Each Region. At the acceleration zone, the desired feedrate v_m is obtained as follows:

$$v_m = v_{\text{start}} + (1 + k) J_1 T_s^2 n_a^2. \tag{1}$$

By imposing the jerk limits values on J_1 and maximum allowable acceleration A, n_a is obtained as follows:

$$n_a = \left[\sqrt{\frac{v_m - v_{\text{start}}}{(1 + k) J_1 T_s^2}} + 1 \right], \quad n_a \geq \frac{A}{J_1 T_s}. \tag{2}$$

If there is $n_a < A/J_1 T_s$, we can modify k to make $n_a \geq A/J_1 T_s$; That is,

$$k_1 = \left[\frac{J_1 (v_m - v_{\text{start}})}{A^2} - 2 \right]. \tag{3}$$

Similarly, at deceleration zone, n_b can be obtained as follows:

$$v_m = v_{\text{end}} + (1 + k) J_2 T_s^2 n_b^2, \tag{4}$$

$$n_b = \left[\sqrt{\frac{v_m - v_{\text{end}}}{(1 + k) J_2 T_s^2}} + 1 \right], \quad n_b \geq \frac{D}{J_2 T_s}. \tag{5}$$

If $n_b < D/J_2 T_s$, then

$$k_2 = \left[\frac{J_2 (v_m - v_{\text{end}})}{D^2} - 2 \right], \tag{6}$$

$$k = \min (k_1, k_2). \tag{7}$$

n_a and n_b can be obtained by (2) and (5), respectively. If k is changed, n_a and n_b are calculated by k determined by (7).

Given the numbers of interpolation steps for all regions (except the constant-speed region) specified, the travel length L is obtained as follows:

$$\begin{aligned} 2L = {} & (2 + k) \left((v_{\text{start}} + v_m) n_a + (v_{\text{end}} + v_m) n_b \right) T_s \\ & + 2 v_m n_c T_s. \end{aligned} \tag{8}$$

Then the number of interpolation step n_c for constant-speed region is calculated as follows:

$$n_c = \left[\frac{2L - (2 + k) \left((v_{\text{start}} + v_m) n_a + (v_{\text{end}} + v_m) n_b \right) T_s}{2 v_m T_s} \right] + 1. \tag{9}$$

The numbers of interpolation steps for different regions are expressed as

$$\begin{aligned} & N_1 = n_a, \quad N_2 = k n_a, \quad N_3 = n_a, \quad N_4 = n_c, \\ & N_5 = n_b, \quad N_6 = k n_b, \quad N_7 = n_b. \end{aligned} \tag{10}$$

n_c is rounding. There are three ways to eliminate rounding error. One is to change the desired feedrate v_m. The other is to alter starting point feedrate v_{start}. The third is to modify the end point feedrate v_{end}.

2.1.2. Modifying the Desired Feedrate. If the feedrate v_m is modified to v'_m, it can be obtained from (8), as follows:

$$v'_m = \frac{2L - (2+k)\left(n_a v_{\text{start}} + n_b v_{\text{end}}\right) T_s}{\left(2n_c + (2+k)\left(n_a + n_b\right)\right) T_s}. \tag{11}$$

The jerk at acceleration-speed zone can be recalculated from (1):

$$J'_1 = \frac{v'_m - v_{\text{start}}}{(1+k)\, n_a^2 T_s^2}. \tag{12}$$

Similarly, the jerk at deceleration-speed zone can be recomputed as follows:

$$J'_2 = \frac{v'_m - v_{\text{end}}}{(1+k)\, n_b^2 T_s^2}. \tag{13}$$

The readjusted values of jerks and feedrate should not be greater than the maximum allowable values. Considering (9), (11), (12), and (13), the following in-equations are valid:

$$J'_1 \le J_1, \qquad J'_2 \le J_2, \qquad v'_m \le v_m. \tag{14}$$

For the machining, the feedrate cannot be changed too drastically. Therefore, it is necessary to check the feedrate variation η:

$$\eta = \left| \frac{v_m - v'_m}{v_m} \right|. \tag{15}$$

Then we can arrived at

$$\eta < \frac{1}{n_a + n_b + n_c}. \tag{16}$$

For normal block, the value of $n_a + n_b + n_c$ is much greater than 1, so that the feedrate variation can be neglected.

The S-curve acc/dec sampled-data formulation for each region can be determined by the method stated in [15]. It is

$$f(t) = \begin{cases} v_{\text{start}} + pJ'_1\left(3n^2 + 3n + 1\right), & 0 \le n < n_a, \\ v_{\text{start}} + 3pJ'_1 n_a\left(2n_2 + n_a + 1\right), & 0 \le n_2 < kn_a, \\ v_{\text{start}} + pJ'_1\left(3n_a^2 + 6kn_a^2 + 6n_a n_3 - 3n_3^2 - 3n_3 - 1\right), & 0 \le n_3 < n_a, \\ v'_m, & 0 \le n_4 < n_c, \\ v_{\text{end}} + pJ'_2\left((6+k)\, n_b^2 - 3\left(n_5 - (k+2)\, n_a\right)^2 \right. & \\ \quad \left. -3\left(n_5 - (k+2)\, n_a\right) - 1\right), & 0 \le n_5 < n_b, \\ v_{\text{end}} + 3pJ'_2 n_b\left(n_a + 2n_b - 2n_6 - 1\right), & 0 \le n_6 < kn_b, \\ v_{\text{end}} + pJ'_2\left(3n_b^2 - 6n_b n_7 - 3n_b + 3n_7^2 + 3n_7 + 1\right), & 0 \le n_7 < n_b, \end{cases} \tag{17}$$

where

$$\begin{gathered} p = \frac{1}{6}T_s^2, \qquad n_2 = n - n_a, \qquad n_3 = n - (k+1)\, n_a, \\ n_4 = n - (k+2)\, n_a, \qquad n_5 = n - (k+2)\, n_a - n_c, \\ n_6 = n - (k+2)\, n_a - n_c - n_b, \\ n_7 = n - (k+2)\, n_a - n_c - (k+1)\, n_b. \end{gathered} \tag{18}$$

2.1.3. Altering the Starting Point Feedrate. If the starting point feedrate is changed, n_c is obtained as follows:

$$n_c = \left[\frac{2L - (2+k)\left(\left(v_{\text{start}} + v_m\right) n_a + \left(v_{\text{end}} + v_m\right) n_b\right) T_s}{2v_m T_s} \right]. \tag{19}$$

The starting point feedrate can be calculated from (8) as follows:

$$v'_{\text{start}} = \frac{\left(v_{\text{end}} + v_m\right) n_b}{n_a} + \frac{2L - 2v_m n_c T_s}{(2+k)\, n_a T_s} - v_m. \tag{20}$$

The jerk at acceleration-speed zone can be recomputed from (1):

$$J'_1 = \frac{v_m - v'_{\text{start}}}{(1+k)\, n_a^2 T_s^2}. \tag{21}$$

The sampled-data formulations of S-curve acc/dec for each region are similar to (13). Considering (19), (20), and (21), there is a relationship as follows:

$$J'_1 \le J_1, \qquad v'_m = v_m, \qquad v'_{\text{start}} \ge v_{\text{start}}. \tag{22}$$

2.1.4. Readjusting the End Point Feedrate. If the end point feedrate is modified, n_c is obtained from (19). The end point feedrate is recalculated from (8) as follows:

$$v'_{\text{end}} = \frac{2L - 2v_m n_c T_s}{(2+k)\, n_b T_s} + \frac{\left(v_{\text{start}} + v_m\right) n_a}{n_b} - v_m. \tag{23}$$

The jerk at deceleration-speed zone can be derived from (3):

$$J'_2 = \frac{v_m - v'_{\text{end}}}{(1+k)\, n_b^2 T_s^2}. \tag{24}$$

The sampled-data formulations of S-curve acc/dec for each region are similar to (13). Considering (19), (23), and (24), we can reach

$$J'_2 \le J_2, \qquad v'_m = v_m, \qquad v'_{\text{end}} \ge v_{\text{end}}. \tag{25}$$

We can adjust v_{start}, v_{end}, v_m, J_1, or J_2 to discretizing linear/circular path with S-curve acc/dec. This provides a way of calculating the trajectory machining parameters for look-ahead interpolation.

2.2. Feedrate and Travel Length Constraints. In order to generate a feedrate profile, it is necessary to check the path which is a normal or short block. The normal block has acceleration-speed zone, constant-speed zone, and deceleration-speed zone, shown in Figure 1. The short block does not have constant-speed zone. The S-curve acc/dec sampled-data formulations of normal block are (17). The formulations of short block are similar to (17). According to the travel length, the path is divided into seven types. The normal block refers to Type 1. Types 2 to 7 belong to the short block. Type 1 should satisfy

$$L \geq 0.5\,(2+k)\left(\left(v_{\text{start}}+v_m\right)n_a+\left(v_{\text{end}}+v_m\right)n_b\right)T_s. \quad (26)$$

Type 2 which has no constant-speed zone (that means n_c is zero) is shown in Figure 2(a). It has an acceleration-speed zone and a deceleration-speed zone with constant-acceleration regions. It accelerates to a certain speed from starting point feedrate, then changes to deceleration-speed zone and reaches the end point at feedrate v_{end}. With the travel length shortening, the time of constant-acceleration region is decreasing. In case of Type 2, the following equation should be satisfied:

$$\begin{gathered} L < 0.5\,(2+k)\left(\left(v_{\text{start}}+v_m\right)n_a+\left(v_{\text{end}}+v_m\right)n_b\right)T_s, \\ L \geq \left(\left(v_{\text{start}}+v_m\right)n_a+\left(v_{\text{end}}+v_m\right)n_b\right)T_s. \end{gathered} \quad (27)$$

Type 3 has neither constant-speed zone nor constant-acceleration regions (that means n_b and n_c are zero), as shown in Figure 2(b). Type 3 also has acceleration-speed zone and deceleration-speed zone. The difference between Type 3 and Type 2 is that Type 3 has no constant-acceleration regions. The condition in Type 3 is represented as follows:

$$\begin{gathered} v_{\text{start}} < v_{\text{end}} < v_m, \\ L < \left(\left(v_{\text{start}}+v_m\right)n_a+\left(v_{\text{end}}+v_m\right)n_b\right)T_s, \\ 2L \geq \left(\left(2+k\right)n_a\left(v_{\text{start}}+v_m\right)\right)T_s \end{gathered} \quad (28)$$

or

$$\begin{gathered} v_{\text{end}} < v_{\text{start}} < v_m, \\ L < \left(\left(v_{\text{start}}+v_m\right)n_a+\left(v_{\text{end}}+v_m\right)n_b\right)T_s, \\ 2L \geq \left(\left(2+k\right)n_b\left(v_{\text{start}}+v_m\right)\right)T_s. \end{gathered} \quad (29)$$

In case end point feedrate is equal to v_m, Type 4 that has constant-speed and acceleration-speed zones is to accelerate to speed v_m and keeps moving at this speed until end point is reached. Or if starting point feedrate is equal to v_m, Type 4 that has constant-speed and deceleration-speed zones keeps moving at speed v_m until the deceleration point is reached; then the speed decelerates to v_{end}. Equation (29) needs to be satisfied for Type 4, as shown in Figure 2(c):

$$\begin{gathered} v_{\text{start}} < v_{\text{end}} = v_m, \\ 2L \geq (2+k)\,n_a\left(v_{\text{start}}+v_{\text{end}}\right)T_s \\ \text{or} \\ v_{\text{end}} < v_{\text{start}} = v_m, \\ 2L \geq (2+k)\,n_b\left(v_{\text{start}}+v_{\text{end}}\right)T_s. \end{gathered} \quad (30)$$

The difference between Type 5 and Type 4 is that the starting and end point feedrates of Type 5 are not equal to v_m, as shown in Figure 2(d). If end point feedrate v_{end} is greater than starting point feedrate v_{start}, Type 5 accelerates to end point feedrate v_{end} from starting point feedrate v_{start}. If end point feedrate v_{end} is less than starting point feedrate v_{start}, Type 5 decelerates to end point feedrate v_{end} from v_{start}. In case of Type 5, (30) should be satisfied:

$$\begin{gathered} v_{\text{start}} < v_{\text{end}} < v_m, \\ 2L < (2+k)\,n_a\left(v_{\text{start}}+v_m\right)T_s, \\ L \geq n_a\left(v_{\text{start}}+v_{\text{end}}\right)T_s \\ \text{or} \\ v_{\text{end}} < v_{\text{start}} < v_m, \\ 2L < (2+k)\,n_a\left(v_{\text{start}}+v_m\right)T_s, \\ L \geq n_b\left(v_{\text{start}}+v_{\text{end}}\right)T_s. \end{gathered} \quad (31)$$

Without constant-acceleration region, Type 6 has only acceleration-speed or deceleration-speed zone, shown in Figure 2(e). It is checked as

$$\begin{gathered} v_{\text{start}} < v_{\text{end}} \leq v_m, \\ L < n_a\left(v_{\text{start}}+v_{\text{end}}\right)T_s, \\ L \geq \left(v_{\text{start}}+v_{\text{end}}\right)T_s \\ \text{or} \\ v_{\text{end}} < v_{\text{start}} \leq v_m, \\ L < n_b\left(v_{\text{start}}+v_{\text{end}}\right)T_s, \\ L \geq \left(v_{\text{start}}+v_{\text{end}}\right)T_s. \end{gathered} \quad (32)$$

If the travel length is short enough, the type of S-curve acc/dec refers to Type 7, shown in Figure 2(f). Starting with

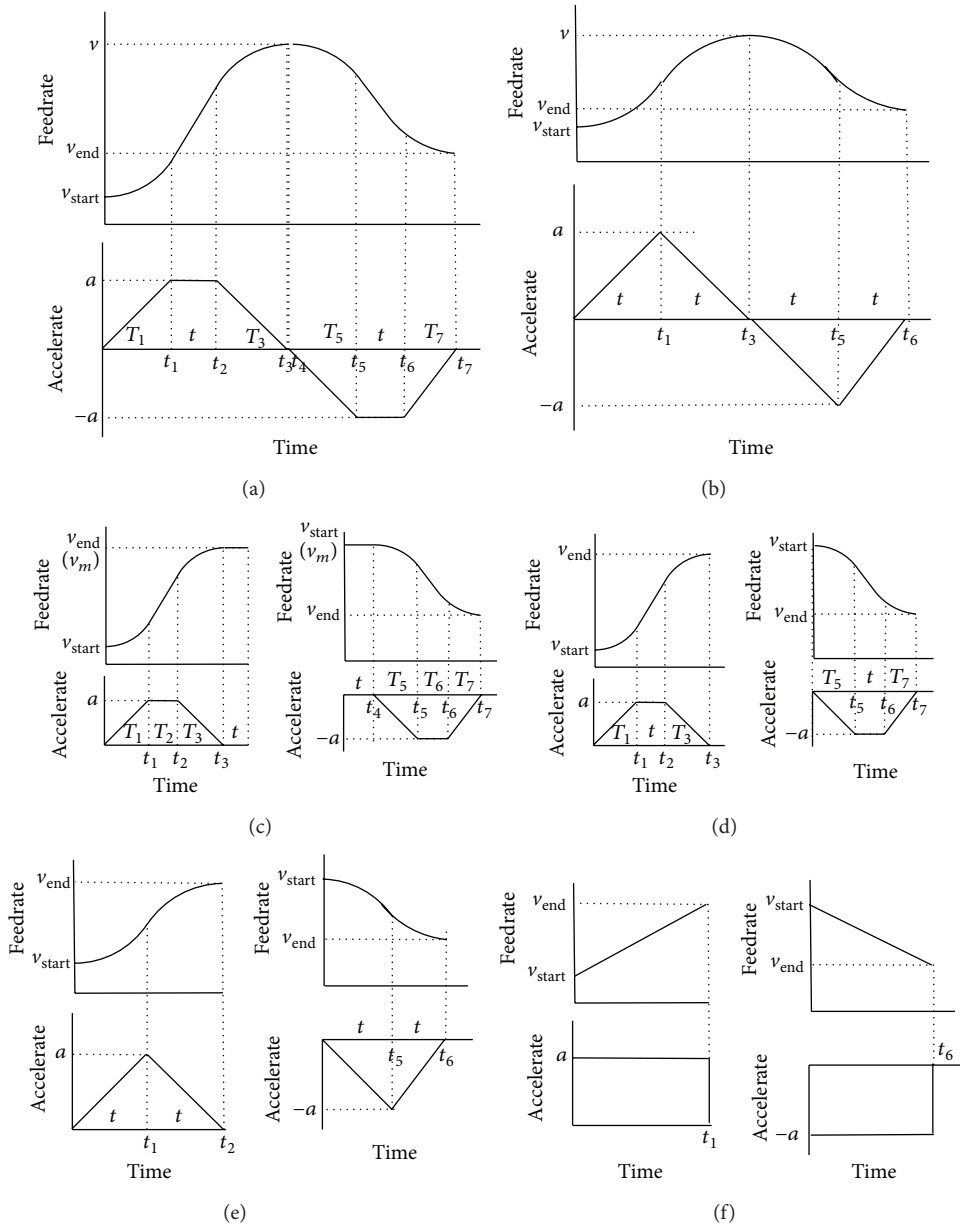


FIGURE 2: Various kinematics profiles for movement of specified travel length.

v_{start}, Type 7 accelerates or decelerates to v_{end} at the end point. Type 7 is tested as

$$\begin{aligned} v_{\text{start}} < v_{\text{end}} \le v_m, \\ L < \left(v_{\text{start}} + v_{\text{end}}\right) T_s \\ \text{or} \\ v_{\text{end}} < v_{\text{start}} \le v_m, \\ L < \left(v_{\text{start}} + v_{\text{end}}\right) T_s. \end{aligned} \tag{33}$$

2.3. Corner Angle Constraints. The feedrate and its direction could be changed at the path turning point, which may cause acceleration. The acceleration that should be less than the maximum allowable acceleration $a_{\max}$ is associated with the turning angle α and the feedrate v_t at the corner point. v_t should meet the following [11]:

$$v_t \le \frac{a_{\max} \cdot T_s}{2 \sin\left(\alpha/2\right)}, \tag{34}$$

$$v = \min\left(v_t, v_m\right). \tag{35}$$

The corner angle is the angle between the path tangent τ_i of forward direction and the next path tangent τ_{i+1} at the turning point, shown in Figure 3. There are four different types at the turning point: line to line, line to arc, arc to line,

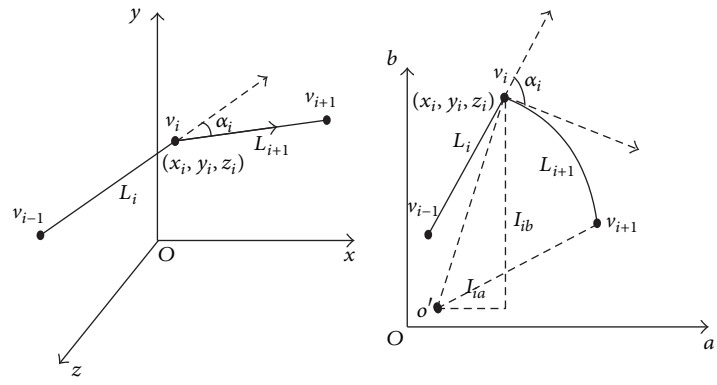


Figure 3: The turning angle.

and arc to arc. The corner angle α between two consecutive path segments is given as

$$\cos\alpha = \frac{\tau_i \cdot \tau_{i+1}}{|\tau_i|\,|\tau_{i+1}|}. \tag{36}$$

3. Look-Ahead Scheme

The look-ahead algorithm which takes the characteristics of path, machine tool, and feedrate profile into consideration realizes the smooth transfer in interpath. The look-ahead scheme assures the smooth movement and the accessibility of the starting point and end point feedrates of each path.

The interpolation of G01, G02, and G03 codes can be achieved by adjusting starting point feedrate v_{start}, end point feedrate v_{end}, and feedrate v_m according to the travel length and the machining parameters. There are three types of algorithm to adjust machining parameters.

Algorithm A. The interpolation can be achieved by modifying the starting point feedrate v_{start} and jerk J_1.

Algorithm B. The interpolation can be completed by adjusting the end point feedrate v_{end} and jerk J_2.

Algorithm C. The interpolation can be realized by recalculating the feedrate v_m and jerks J_1, J_2, but the starting point and end point feedrates are not changed.

3.1. Algorithm of Look-Ahead Scheme. In order to guarantee the accessibility of the starting point and end point feedrates of each path, the look-ahead scheme is realized as follows.

Step 1 (calculating turning point feedrates). The controller reads j paths before interpolation and the look-ahead path number j is determined by the controller. The starting point feedrate of the first path and the end point feedrate of last path are v^0_{start} and v^0_{end}, respectively. The feedrates at turning point are obtained by (35).

Step 2 (dividing the paths into blocks). Starting with the first path, the end point feedrates of each path are determined by Algorithm B in order until the last path. The break paths are checked and recorded. The break path whose travel length is so short that the path cannot be interpolated with the S-curve acc/dec is referred to as Type 7. The starting point and end point feedrates of the break path are set according to Type 7.

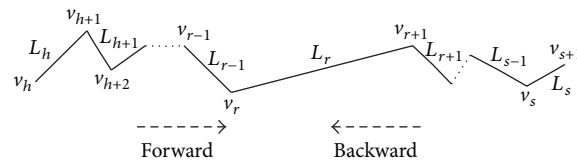


Figure 4: The look-ahead path.

The paths are divided into blocks by break paths. All the paths lying in every two adjacent break paths are classified into one block. The starting point feedrate of the block is the end point feedrate of the front break path. The end point feedrate of the block is the starting point feedrate of the behind break path.

The longest perfect paths of each block are checked and recorded. The perfect path that can be determined by Algorithm C is referred to as the path in which the interpolation can be completed by adjusting feedrate v_m and jerks J_1, J_2; the starting and end point feedrates are not changed. Types 1 to 6 belong to the perfect path.

Step 3 (determine the starting and end point feedrates). Supposing paths $L_h, L_{h-1}, \ldots, L_s$ belong to a block, L_r is the longest perfect path, shown in Figure 4. Starting with the path L_h, the end point feedrates v_g $(g = h{+}1, h{+}2, h{+}3, \ldots, r{-}1, r)$ of paths are determined by Algorithm B in order until the path L_{r-1}. The end point feedrate of path L_{r-1} and the starting point feedrate of the path L_r are v_r. Starting with L_s, the starting point feedrate v_w $(w = s + 1, s, \ldots, r + 1)$ of path is determined by Algorithm A in the reverse order until the path L_{r+1}. The starting point feedrate of path L_{r+1} and the end point feedrate of path L_r are v_{r+1}. The starting point and end point feedrates of the longest perfect path L_r are v_r and v_{r+1}, respectively. Since L_r is perfect path, it can be interpolated by recalculating the feedrate v_m and jerks J_1, J_2, the starting point and end point feedrates are not changed. Step 3 is repeated until all the blocks are treated and all the starting and end

Table 1: Coordinates of path to be machined (mm).

Number	Type	Coordinate value	Center of arc	Number	Type	Coordinate value	Center of arc
0		(2.16, 10.672)					
1	Line	(3.952, 11.016)		6	Circle	(12.416, 8.4)	(11.432, 87.649)
2	Line	(5.768, 11.064)		7	Line	(18.552, 11.016)	
3	Line	(7.568, 10.808)		8	Line	(19.696, 4.48)	
4	Line	(9.312, 10.272)		9	Line	(20.936, 5.776)	
5	Line	(10.936, 9.46)		10	Line	(22.256, 5.232)	

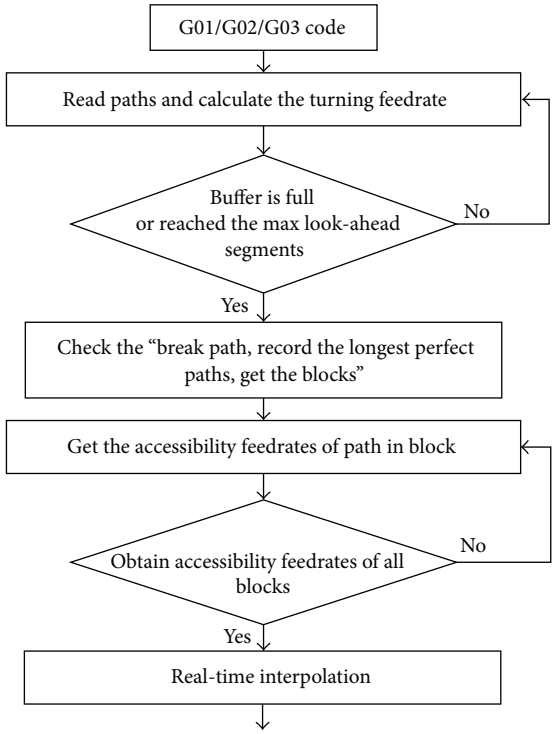


Figure 5: Flowchart of look-ahead scheme.

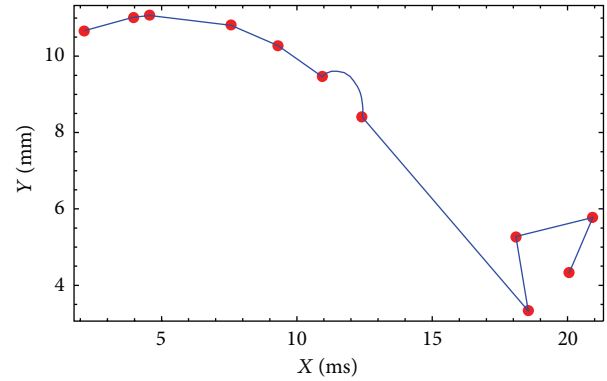


Figure 6: The tool paths of the look-ahead scheme.

point feedrates of paths are recorded in the look-ahead buffer queue.

Step 4 (real-time interpolation). Take the first path in the look-ahead buffer queue. Calculate the time of every region with the S-curve acc/dec and compute the interpolation point in S-curve acc/dec sampled-data formulations. The movements of mechanics are executed according to the interpolation points. When the interpolation of current line is completed, Step 4 is repeated until all the paths are finished. The look-ahead scheme flowchart is shown in Figure 5.

4. Simulation and Experiment

Several simulations and experiments applying the proposed approach were made on a CNC milling machine, in which $v_m = 2$ m/min, $A = D = 40$ m/s^2, $J_1 = J_2 = 10$ m/s^3, and $T_s = 1$ ms.

Table 1 shows the coordinates of the starting and end points of 10 small linear/circular segments describing a piece of 2D curve which is deliberately chosen to test the proposed look-ahead scheme. As shown in Figure 6, the first five paths are lines. The sixth path is a circle. There are two sharp corners from the seventh to the tenth path.

The feedrate profiles are shown in Figure 7. Without look-ahead scheme, it has always the acceleration and deceleration phases in each path and the machine tool never reaches the desired feedrate, as shown in Figure 7(a). With the proposed look-ahead scheme, although there are two sharp corners, the feedrate accelerates to desired feedrate and almost keeps moving in this speed until the end point, as shown in Figure 7(b).

The proposed look-ahead scheme can decrease the numbers of the break paths and zero feedrate. As shown in the first row of Table 2, the travel length of path 2 is too short. In [12–14], path 2 is a break path whose starting and end point feedrates are zero.

With this look-ahead scheme, path 2 is not a break path, as shown in Table 2. As shown in Figure 5, there are two sharp corners at the turning points of point 8 and 10. In [12–14], the end point feedrates of seventh and ninth path segment are zero. With this proposed look-ahead scheme, the end point feedrates of seventh and ninth path segment are not zero, as shown in Table 2. As a result, the machining time with the proposed look-ahead scheme was approximately 1 s, while the

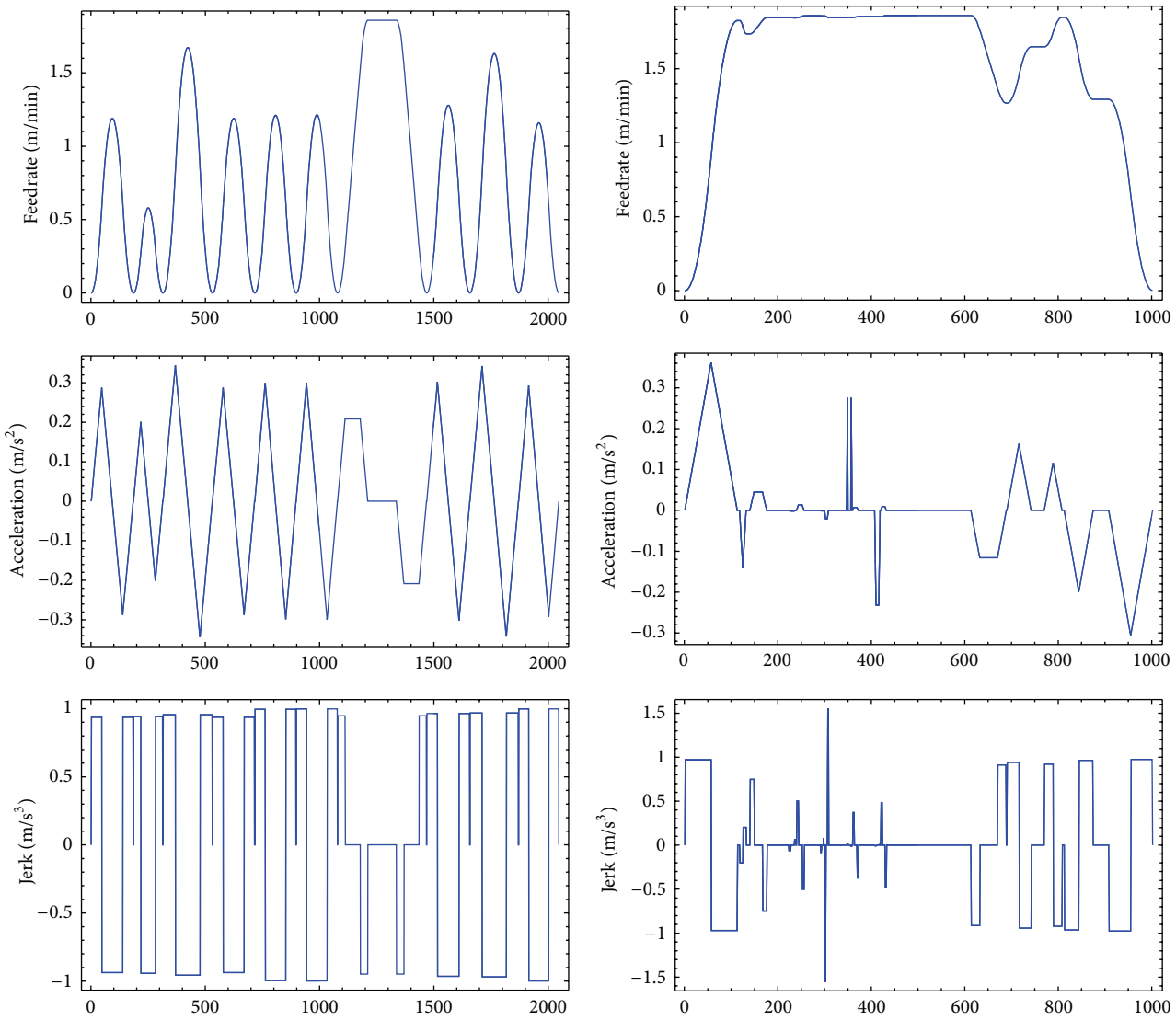


Figure 7: Feedrate profiles of S-curve acc/dec characteristics without/with look-ahead.

Table 2: The end point feedrate (m/min) and travel length (mm).

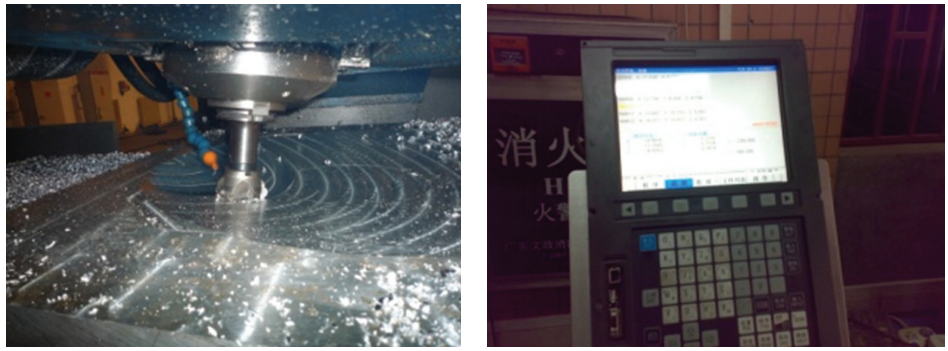
Number	1	2	3	4	5	6	7	8	9	10
End feedrate	1.82	1.74	1.84	1.86	1.85	1.85	1.27	1.65	1.29	0
Travel length	1.825	0.617	3.011	1.825	1.816	1.82	7.96	2.0	2.883	1.701

time without the look-ahead scheme was about 2 s, as shown in Figure 7.

The part shown in Figure 8 was machined on a five-axis machining center controller with an in-house developed CNC running on microC/OS-II real-time operation system. The experimental results showed that the machining time was dramatically shortened when the proposed approach is applied. Moreover, the feedrates are smoother than those without look-ahead approach and the final contour accuracy of the parts is satisfying.

5. Conclusion

Look-ahead scheme is the key issue of high speed and high precision CNC machining for small paths. We propose a feedrate profile generation approach and corresponding look-ahead algorithm with whole S-curve acc/dec. The proposed look-ahead scheme can handle both small line and circle segments. The feedrate profiles generation effectively predetermines the numbers of interpolation steps for different regions and the discretion feedrate of the path with S-curve acceleration and deceleration. As a result, the look-ahead algorithms with S-curve acceleration and deceleration easily meet real-time requirements. The simulations and



(a) Machined workpiece (b) In-house CNC

Figure 8: A machined workpiece with the developed look-ahead approach.

experiments showed that the proposed approach is feasible and effective.

Acknowledgments

The work presented in this paper was supported by the Major National S&T Program of China (Grant no. 2011ZX04002-151) and National 863 Program (Grant no. 2011AA04A104).

References

[1] D. C. H. Yang and T. Kong, "Parametric interpolator versus linear interpolator for precision CNC machining," *Computer-Aided Design*, vol. 26, no. 3, pp. 225–234, 1994.

[2] M. Shpitalni, Y. Koren, and C. C. Lo, "Realtime curve interpolators," *Computer-Aided Design*, vol. 26, no. 11, pp. 832–838, 1994.

[3] H. Yau and M. Kuo, "NURBS machining and feed rate adjustment for high-speed cutting of complex sculptured surfaces," *International Journal of Production Research*, vol. 39, no. 1, pp. 21–41, 2001.

[4] M. Zhang, W. Yan, C. Yuan, D. Wang, and X. Gao, "Curve fitting and optimal interpolation on CNC machines based on quadratic B-splines," *Science China Information Sciences*, vol. 53, no. 1, pp. 1–12, 2010.

[5] H.-T. Yau and J.-B. Wang, "Fast Bezier interpolator with real-time lookahead function for high-accuracy machining," *International Journal of Machine Tools and Manufacture*, vol. 47, no. 10, pp. 1518–1529, 2007.

[6] J. B. Wang and H. T. Yau, "Real-time NURBS interpolator: application to short linear segments," *International Journal of Advanced Manufacturing Technology*, vol. 41, no. 11-12, pp. 1169–1185, 2009.

[7] J. A. Tao, S. L. Gao, Q. N. You, and Q. Shi, "Lookahead and acceleration/deceleration algorithms for micro-line blocks machining," *Journal of Computer-Aided Design and Computer Graphics*, vol. 22, no. 9, pp. 1570–1586, 2010.

[8] L. X. Zhang, R. Y. Sun, X. S. Gao, and H. B. Li, "High speed interpolation for micro-line trajectory and adaptive real-time look-ahead scheme in CNC machining," *Science China Technological Sciences*, vol. 54, no. 6, pp. 1481–1495, 2011.

[9] L. B. Zhang, Y. P. You, J. He, and X. F. Yang, "The transition algorithm based on parametric spline curve for high-speed machining of continuous short line segments," *International Journal of Advanced Manufacturing Technology*, vol. 52, no. 1–4, pp. 245–254, 2011.

[10] C. Lee, "Generation of velocity profiles with speed limit of each axis for high-speed machining using look-ahead buffer," *International Journal of Precision Engineering and Manufacturing*, vol. 11, no. 2, pp. 201–208, 2010.

[11] J. Hu, L. Xiao, Y. Wang, and Z. Wu, "An optimal feedrate model and solution algorithm for a high-speed machine of small line blocks with look-ahead," *International Journal of Advanced Manufacturing Technology*, vol. 28, no. 9, pp. 930–935, 2006.

[12] Y. Wang, L. Xiao, S. Zeng, Z. Wu, and S. Zhong, "Optimal feedrate model and solution for high-speed machining of small line blocks with look-ahead," *Journal of Shanghai Jiaotong University*, vol. 38, no. 6, pp. 901–904, 2004.

[13] J. He, Y. You, H. Chen, and H. Wang, "A fast nested look-ahead algorithm with S-shape acceleration and deceleration," *Acta Aeronautica et Astronautica Sinica*, vol. 31, no. 4, pp. 842–851, 2010.

[14] Y. N. Cao, T. M. Wang, Y. D. Chen, and H. X. Wei, "Application of pre-interpolation S-shape acceleration/deceleration in CNC look-ahead interpolation algorithm," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 33, no. 5, pp. 594–599, 2007.

[15] Y. Chen, H. Wei, K. Sun, M. Liu, and T. Wang, "Algorithm for smooth S-curve feedrate profiling generation," *Chinese Journal of Mechanical Engineering*, vol. 24, no. 2, pp. 237–247, 2011.

Hindawi Publishing Corporation
Advances in Mechanical Engineering
Volume 2013, Article ID 601612, 11 pages
http://dx.doi.org/10.1155/2013/601612

Research Article

A Scale Adaptive Mean-Shift Tracking Algorithm for Robot Vision

Yimei Kang,[1,2] Wandong Xie,[1,3] and Bin Hu[1]

[1] *College of Software, Beihang University, Beijing 100191, China*
[2] *The Key Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[3] *Information Centre of China North Group Corporation, Beijing 100089, China*

Correspondence should be addressed to Yimei Kang; 2368103448@qq.com

Received 22 March 2013; Accepted 15 May 2013

Academic Editor: Hongxing Wei

The Mean-Shift (MS) tracking algorithm is an efficient tracking algorithm. However, it does not work very well when the scale of a tracking target changes, or targets are occluded in the movements. In this paper, we propose a scale-adaptive Mean-Shift tracking algorithm (SAMSHIFT) to solve these problems. In SAMSHIFT, the corner matching is employed to calculate the affine structure between adjacent frames. The scaling factors are obtained based on the affine structure. Three target candidates, generated by the affine transformation, the Mean Shift and the Mean Shift with resizing by the scaling factors, respectively, are applied in each iteration to improve the tracking performance. By selecting the best candidate among the three, we can effectively improve the scale adaption and the robustness to occlusion. We have evaluated our algorithm in a PC and a mobile robot. The experimental results show that SAMSHIFT is well adaptive to scale changing and robust to partial occlusion, and the tracking speed is fast enough for real-time tracking applications in robot vision.

1. Introduction

Real-time target tracking is a critical task in robot vision. For example, a service robot needs to track a target person to provide services; a surveillance robot may be required to track a suspected object to get more information. In such real-time target tracking applications, the scale change and occlusion of a target are the most challenging issues besides the time performance. Among various tracking algorithms, the Mean-Shift tracking [1] is one of the most efficient tracking algorithms for real-time applications. But the Mean-Shift tracking algorithm has poor performance when the scale change of a target occurs or targets are occluded. In this paper, we focus on solving these problems of the Mean-Shift tracking algorithm for real-time target tracking in robot vision.

Many techniques have been proposed to solve these problems in the previous work. In [1], a method is proposed to reduce the interference of background near the target area in Mean-Shift tracking by using the background-weighted histogram (BWH). In this method, the radius of the kernel is modified by ±10%. The Mean-Shift tracking algorithm is independently applied three times with different kernel scales. The scale which yields the largest Bhattacharyya coefficient is selected as the kernel scale. Ning et al. [2] proved that BWH is equivalent to the common target representation. No new information in BWH is introduced to improve the performance of Mean-Shift tracking. Then a corrected BWH (CBWH) was proposed by transforming the original target model but not the target candidates. The CBWH scheme significantly reduces the background interference in a target area by enhancing prominent features of a target model and reducing the impact of similar image features shared by the target and background. In [3], an algorithm is proposed based on CBWH. The algorithm performs well when targets become smaller. However, when targets become larger, especially when the scale change exceeds the size of tracking window, it cannot work very well as the tracking window shrinks because the Bhattacharyya coefficient always

converges to the local maximum value in a smaller searching window [4].

An automatic bandwidth selection method is proposed in [5]. This method applies backward tracking and object centroid registering to improve the performance of Mean-Shift tracking. This method is based on the theorem that the changes of the target scale and target position within the kernel will not impact the tracking accuracy of the Mean-Shift tracking algorithm. However, this method cannot be employed when the scale change of a target becomes smaller. In [6], a method called CAMSHIFT is proposed to address this problem based on invariant moments. This method is not suitable for real-time tracking due to the complexity of the calculation of the second order central moment. Hu et al. [7] presented an enhanced Mean-Shift tracking method similar to CAMSHIFT algorithm by using joint spatial-color feature space and a novel similarity measure function. Corresponding eigenvalues are used to monitor the scale of the object. These algorithms have great deviation in tracking and might lose the target when the target is occluded. Affine projection is also introduced to solve the problem of scale adaptation [8–10]. However, simply introducing affine transformation in these methods cannot get a satisfied scale adaptation. The interest point matching in these methods is not efficient to obtain correct affine structures. Furthermore, tracking failure occurs when the target is occluded.

In this paper, we proposed a scale-adaptive Mean-Shift tracking algorithm (SAMSHIFT) to solve these problems. In our algorithm, we provide three target candidates that are obtained by an efficient affine transformation based on the corner point matching, the Mean-Shift tracking algorithm, and the Mean-shift algorithm with resizing by the scaling factors calculated from the affine structure, respectively. Then the best candidate among the three is selected in such a way that we can effectively improve the scale adaptation and robustness to occlusion.

We evaluate our algorithm in a PC and a mobile robot, respectively. Our algorithm is compared with several representative methods including the traditional Mean-Shift tracking algorithm [1], the traditional Mean Shift with 10% adaption [1], and the spatial colour Mean-Shift tracking algorithm [7]. The experiments show that our algorithms can effectively improve the scale adaptation and occlusion robustness both in the PC and mobile robot. Furthermore, the tracking speed of our algorithm is fast enough for real-time tracking applications in robot vision.

The rest of this paper is organized as follows: in Section 2, the Mean-Shift tracking algorithm is outlined, and its limitation in scale change is analysed. In Section 3, our scale-adaptive Mean-Shift tracking algorithm is presented. In Section 4, we present and analyse the experimental results. The conclusions are given in Section 5.

2. Background

In this section, we first introduce the Mean-Shift tracking algorithm and then analyse existing problems.

The Mean-Shift tracking algorithm is a semiautomatic tracking algorithm. The target window in the first frame is initialized manually or by other recognition algorithms. The Mean-Shift tracking algorithm is an iterative scheme, in which the RGB colour histogram of the original target in the first frame is iteratively compared with that of the target candidate regions in the following frames. The objective function is to maximize the correlation between two colour histograms.

Let $S = \{x_i\}_{i=1,\ldots,n}$ be a sample of independent and identically distributed random variables drawn from some distribution with an unknown density $f(x)$. The size of the target is $M \times N$ pixels. The kernel density estimation used in the traditional Mean-Shift tracking is defined as follows:

$$f(x) = \frac{1}{M \times N} \sum_{i=1}^{M \times N} k_h \left(\left| \frac{x - x_i}{h} \right|^2 \right), \tag{1}$$

where h is the bandwidth, and $h > 0$.

Target candidates with the same scale are established one by one around the point which is mapped from the target centre in previous frame. Then the Bhattacharyya coefficient is applied to calculate the similarity between each candidate in current frame and the target in previous frame. The Bhattacharyya coefficient is defined as follows:

$$d(y) = \sqrt{1 - \rho(y)}, \tag{2}$$

where

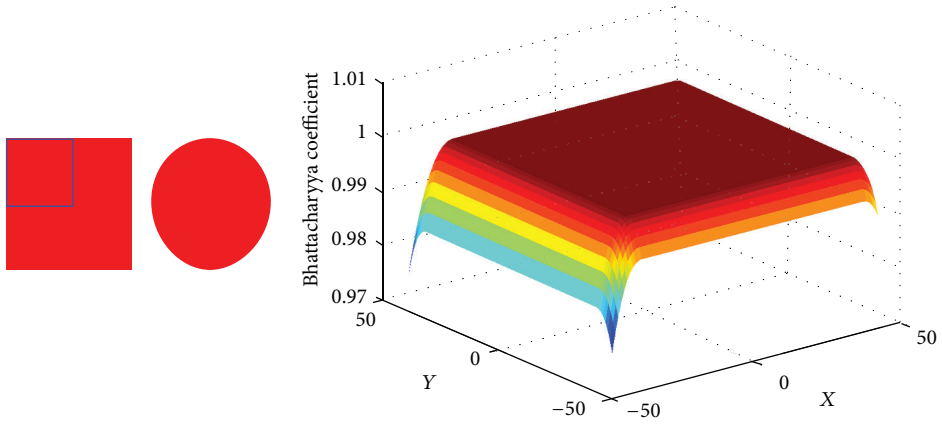
$$\rho(y) \equiv \rho(p(y), q) = \sum_{b=1}^{B} \sqrt{p_b(y)\, q_b}. \tag{3}$$

In (3), q represents the target model, $\rho(y)$ represents the candidate model, and B is the number of bins used to calculate the model.

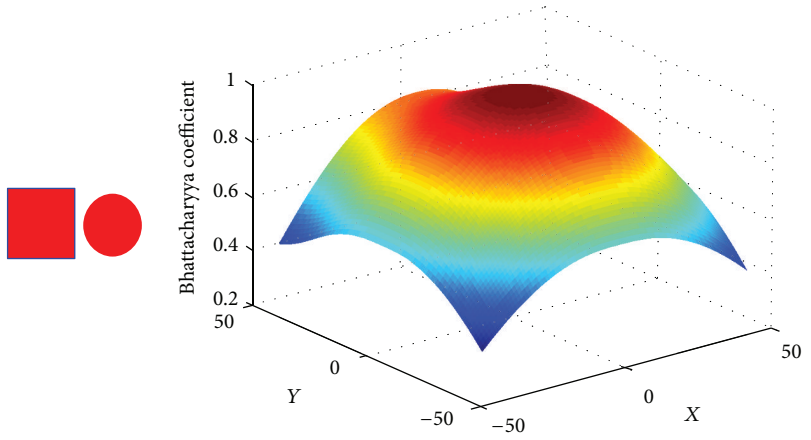
The kernel function $k_h(|(x - x_i)/h|^2)$ is symmetrically centered at point x, so the Mean-Shift tracking algorithm is robust to rotation. The statistical property of the kernel density estimation makes it insensitive to partial occlusion. The kernel bandwidth h plays an important role in the Mean-Shift tracking algorithm. It not only determines the sample weights but also reflects the shape and size of a target. In the Mean-Shift tracking algorithm, the kernel scale is initialized by the first tracking window and fixed through the whole tracking process. Due to its fixed kernel scale, the Mean-Shift tracking algorithm has poor performance when the target scale changes a lot.

Figure 1 shows the performance of the Mean-Shift tracking algorithm when the target scale changes. The red solid square is the target we want to track. The red pie is an object in the background. The blue square frame is the tracking window which is fixed in the Mean-Shift tracking algorithm. The surface charts are the Bhattacharyya coefficient surface between the target and the target candidates. These surface charts are centred within a zone with 45×45 pixels around the true target location.

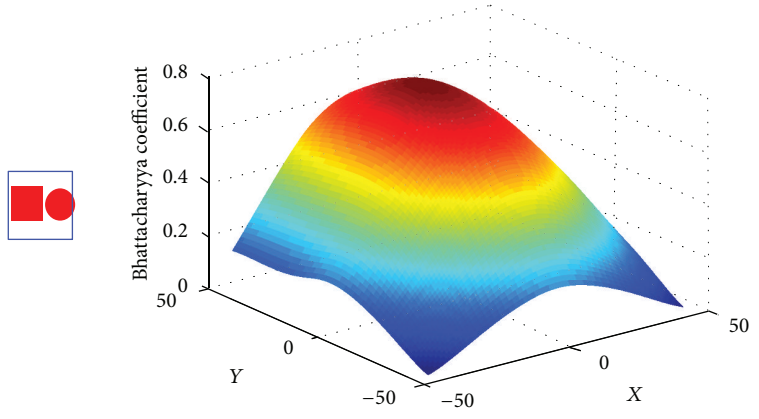
In the Mean Shift tracking algorithm, the deviation generated in each iteration is actually the gradient vector at a certain point on the Bhattacharyya coefficient surface. The process of locating a target is equivalent to the process



(a) The kernel scale is smaller than the target size



(b) The kernel scale equals to the target size



(c) The kernel scale is larger than the target size

Figure 1: The performance analysis of the Mean-Shift algorithm with fixed kernel scales.

of searching the peak along the gradient vector on the Bhattacharyya coefficient surface.

In Figure 1(a), the tracking window is smaller than the target size. The shifts move within a flat zone around the true location. This may lead to locating the target with a great shift because there is no extreme point. In Figure 1(b), the size of the tracking window is the same as the target size. Accordingly there is a single peak near the target location in the Bhattacharyya coefficient surface. This may lead to the accurate target location. As shown in Figure 1(c), a large tracking window will lead the Mean-Shift iteration to converge to an area which is mixed with the target and the background.

Hence, the defaults of fixed kernel scales can be summarized as follows:

(1) When the target becomes smaller, the tracking window contains lots of background information. The background in the tracking window affects the color-histogram distribution seriously. It is prone to cause the tracking centre to shift.

(2) When the target becomes larger, the target will exceed the tracking window. The color-histogram distribution of the tracking window is similar if only the tracking window is within the target area. So the tracking window always drifts within the target area.

3. Proposed Algorithm SAMSHIFT

In this section, we present our scale-adaptive Mean-Shift tracking algorithm called SAMSHIFT. We first provide an overview of SAMSHIFT in Section 3.1 and then introduce its two important functions, Corner Points Matching and Affine Transformation, in Sections 3.2 and 3.3, respectively. Finally, we discuss the robustness to partial occlusion and the time complexity of SAMSHIFT in Section 3.4.

3.1. Overview. Algorithm SAMSHIFT is proposed to improve the scale adaptation and robustness to occlusion of the Mean-Shift tracking algorithm. The targets in adjacent frames in a video have slightly different views for the same object. If we can build a correct model to describe different views including scale changing, then we can more precisely track the target object in videos. In SAMSHIFT, we apply the affine transformation because it is often used to model different views for rotation, transformation and scaling of a rigid object. However, feature point matching is required for affine transformation. We use corner points as interest feature points in SAMSHIFT.

The workflow of SAMSHIFT is illustrated in Figure 2, in which W_i is the Mean-Shift tracking window in the ith frame, C_i is the centre of W_i, and C_i' is the target centre in the ith frame which is mapped from the target centre in the $(i-1)$th frame obtained by the Mean-Shift tracking algorithm. W_i' is W_i resized with s_x and s_y. W_i'' is the tracking window centred at C_i' and resized with s_x and s_y. N_i is the number of matching pairs between W_i and W_{i-1}.

As shown in Figure 2, in SAMSHIFT, we first get a tracking window by the Mean-Shift tracking algorithm. Then

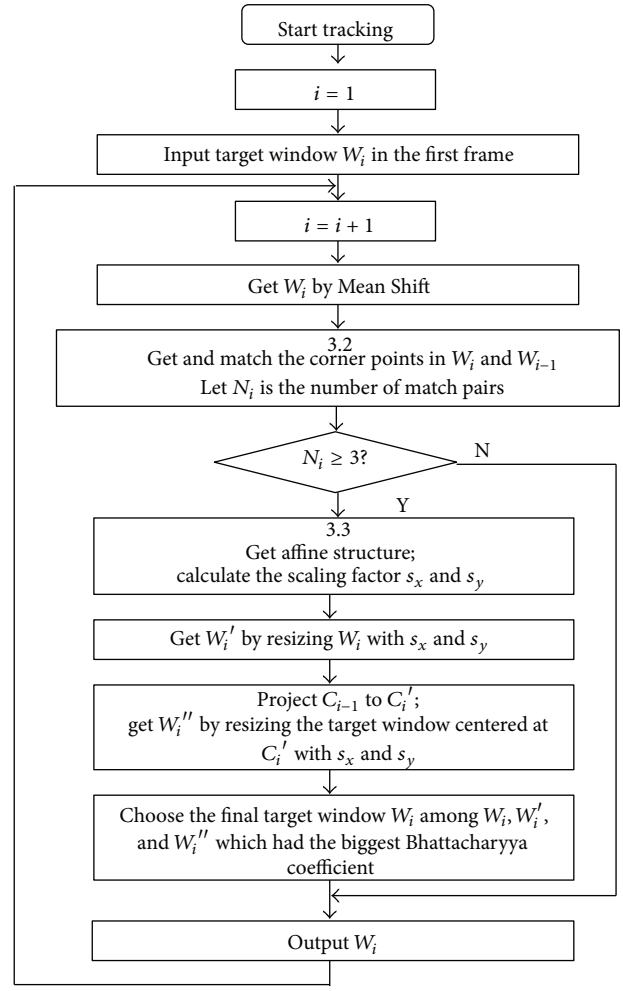


Figure 2: The workflow of SAMSHIFT.

we detect the corner points in the tracking window to match with the corner points in the previous frame (the detailed procedure is described in Section 3.2). Based on these matched corner point pairs, the affine transformation is calculated to get a new target candidate and the scaling factors (the detailed procedure is shown in Section 3.3). The scaling factors are applied to resize the target candidate obtained by the traditional Mean-Shift tracking algorithm. Finally, three target candidates, obtained by the affine transformation, the Mean-Shift tracking algorithm, and the Mean-Shift tracking algorithm with resizing by the scaling factors, respectively, are compared with the target model to select the final target. The candidate which yields the largest Bhattacharyya coefficient is selected as the final target.

3.2. Corner Points Matching. Figure 3 shows an example for the proposed method. The remote control toy car is the target in the video. The first target candidate is obtained by the Mean-Shift tracking algorithm. Then, based on the first candidate, the second target candidate was calculated with the affine transform, and the third candidate is obtained by resizing the first target candidate mapped from a previous

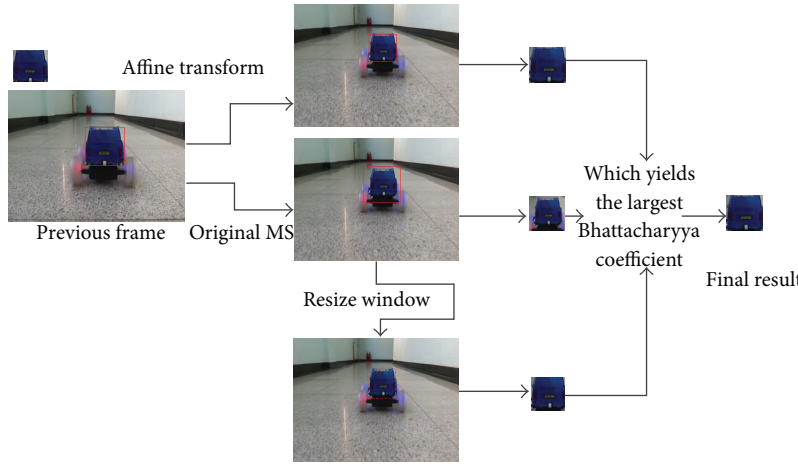


FIGURE 3: An example application of SAMSHIFT.

frame. The final target is chosen among these three candidates by comparing their Bhattacharyya coefficients.

To apply the affine transformation, it is necessary to find interest feature point pairs to calculate the affine structure. In practice, the interest points with some unchanged image features are used to find the relationship between images. False feature point pairs matching may cause incorrect affine structures. Then it will obtain incorrect target scales.

Corner points are formed from two or more edges. The edges are usually used to define the boundary between two different objects or parts of an object. So corner points can be used to calculate the affine structure. In addition, corner points can also be used to distinguish targets from the background with similar colour-histogram distribution. It is known that the Mean-Shift tracking algorithm is based on statistical data comparison, that is, colour-histogram comparison. Therefore, it is sensitive to the difference between the target and the background. However, the target centre calculated by the Mean-Shift algorithm often drifts away from the true location if the background has the similar colour-histogram distribution to the target. Corner points matching is more accurate than such statistic data comparison.

Harris Corner Detector [11] is used in SAMSHIFT because it is suitable for real-time applications. First, corner features are detected in W_i. Suppose I_x and I_y is the grayscale gradients of point $I(x, y)$ at horizontal direction and vertical direction, respectively. Harris Corner Detector is defined as the following:

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}, \tag{4}$$

where

$$\begin{aligned} A &= I_x^2 \otimes \omega, \\ B &= I_y^2 \otimes \omega, \\ C &= I_x I_y \otimes \omega. \end{aligned} \tag{5}$$

In (5), $\otimes$ denotes the convolution operation, ω denotes a Gaussian window centered at $I(x, y)$. we used a 5×5 Gaussian window. A point is a corner point if the corner measure is larger than a threshold. Corner measure R is defined as the following:

$$R = \det(M) - k \times (\text{trace}(M))^2, \tag{6}$$

where $k = 0.04$.

Subsequently, zero-mean normalized cross correlation is employed to seek the best matching pairs among corner features of the ith frame and the $(i-1)$th frame. The similarity between the feature point pairs is defined as follows:

$$\begin{aligned} C\left(x, x'\right) = {} & \frac{1}{(2N+1)^2 \sigma_{I_0}(x)\, \sigma_{I_1}\left(x'\right)} \\ & \times \sum_{i=-N}^{N} \sum_{j=-N}^{N} \left[I_0(x-i, y-j) - \bar{I}_0\right] \\ & \times \left[I_1\left(x'-i, y'-j\right) - \bar{I}_1\right]. \end{aligned} \tag{7}$$

In (7), $\bar{I}_0$ and $\bar{I}_1$ are the average greyscales of the windows with the side of $(2N+1)$ and centred at x and x', respectively. $\sigma_{I_0}(x)$ and $\sigma_{I_1}(x')$ are the standard deviations of the greyscales in the windows centred at x and x', respectively. The value range of $C(x, x')$ is $[-1, 1]$. It is easy to select a unified threshold for all feature points. A feature point pair with the similarity larger than the threshold is considered as a matched pair.

3.3. Affine Transformation. Once the corner points in adjacent frames are detected, the corner matching is used to calculate the affine structure. In 2-dimensional cases, three matched point pairs are required between successive frames to calculate the affine structure [12]. If there are no less than three corner matches, it is possible to determine the

(a) Tracker *without* the orignal MS target candidate

(b) Tracker *with* the orignal MS target candidate

Figure 4: Balance to scale adaptation and occlusion robustness.

affine structure A_i of these matching pairs. According to the principle of affine structure, the interest points satisfy

$$\begin{aligned}\begin{pmatrix} x_i \\ y_i \end{pmatrix} &= A_i \begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix} + b_i \\ &= \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix} + \begin{pmatrix} e_x \\ e_y \end{pmatrix}.\end{aligned} \tag{8}$$

In (8), $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$ and $\begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix}$ reflect positions of the matched points in the ith frame and the $(i-1)$th frame, respectively. A_i is a 2×2 matrix. s_x and s_y are stretch amplitudes in horizontal and vertical directions, respectively. b_i is a 2×1 matrix. e_x and e_y represent the translation in horizontal and vertical directions, respectively.

Since $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$ and $\begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix}$ are known in the above transforming, we can calculate s_x and s_y. In SAMSHIFT, s_x and s_y will be used as the scaling factors in the horizontal and vertical directions, respectively. The target candidate obtained by the Mean Shift will be resized by these scaling factors.

However, mismatches will obtain incorrect affine structure. Hence, RANSAC (RANdom SAmple Consensus) is employed to eliminate the false matching and calculate the final affine structure.

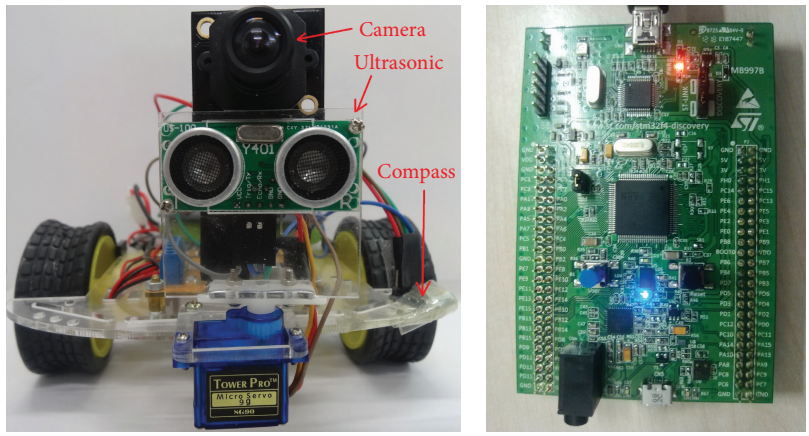
3.4. Discussion. In this section, we discuss the robustness to partial occlusion and the time complexity analysis of the proposed algorithm, respectively.

3.4.1. Robustness to Partial Occlusion. Corner point matching is more accurate than the statistic feature matching. It can reduce the mismatch caused by similar color histogram between the target and background, which may occur in Mean-Shift tracking.

However, corner feature matching on the occluder may obtain incorrect scaling factors. Then the tracking window will be larger or smaller than the true target scale. When the target is occluded, there may be corner matching pairs on the occluder or the background. It will cause the target candidate to be stretched or compressed, or even lost. As shown in Figure 4(a), the moving car is occluded by the tree. The corner feature matches on the tree. The tracking window is compressed.

However, the statistic feature matching in the traditional Mean-Shift tracking algorithm is robust to partial occlusion. Therefore, in SAMSHIFT, the original target candidate obtained by the Mean Shift is also used as one of the target candidates. Together with the resized target candidates, the one which yields the largest Bhattacharyya coefficient is selected as the final target in current frame. By using this scheme, the robustness for partial occlusion of the traditional Mean Shift tracking algorithm is retained. Figure 4(b) shows the tracking performance of our method for the same video. It is robust to partial occlusion and is adaptive to scale changing as well.

3.4.2. Time Complexity Analysis. There are four main components which contribute to the computational cost of SAMSHIFT: the calculation of the gradient information, the Mean Shift tracking, Harris corner detection, and the computation of the affine structure. Let n_h be the pixel number in the tracking window. The time complexity of the proposed algorithm can be expressed as $O(n_h)$. Time cost of



(a) The automatic tracking car (b) The controller board

Figure 5: An automatic target tracking car (ATTC) based on SAMSHIFT.

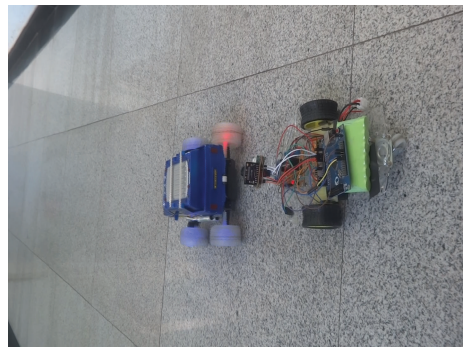


Figure 6: The ATTC is tracking a remote control toy car.

target tracking in each frame is related to the size of the target. The bigger the target is, the more time it will cost.

4. Experiments

In this section, we present and analyse the experimental results. To evaluate the performance of our SAMSHIFT algorithm, we implement it in a PC and a mobile robot. SAMSHIFT is compared with three representative tracking algorithms: the traditional Mean-Shift algorithm, the Mean-Shift algorithm with 10% adaption, and the spatial color Mean-Shift algorithm. Next, we first introduce the experimental environment and the performance metrics and then present and discuss the experimental results.

4.1. Experimental Environments. SAMSHIFT is evaluated in both a PC and a mobile robot. In a PC with Pentium (*R*) 3.0 GHz, we implement the above four algorithms and use a video in VIRAT Ground Video Dataset to test them. In the test video, the image size is converted to 320∗240 pixels.

We also test the four algorithms in a mobile robot. The mobile robot is an automatic target tracking car (ATTC) which is shown in Figure 5(a). STM32F407 ARM+DSP processor is used as the main processor in the controller board as shown in Figure 5(b). The ATTC is equipped with an OV7670 video camera and a compass. The ATTC can be used to track a target in specific environments which are difficult for humans to reach, for example, a very small and narrow space. The tracking algorithms are used to guide the ATTC to track a target. As shown in Figure 6, the target is a remote control toy car.

4.2. Performance Metrics. The following three performance metrics are used to evaluate the four algorithms. First, the tracking results by images are recorded to show the visual tracking effects in the environments with scale changing and occlusion. We also compare the performance of the target centre position errors and relative scale errors. The target centre position error is the relative position error between the centre of the tracking result and that of the true target. Relative scale is scale normalized by the true scale. It is defined as the following:

$$S = \frac{W \times H}{W_t \times H_t}. \tag{9}$$

In (9), W and H are the width and height of the current target scale obtained by a tracking algorithm, respectively. W_t and H_t are the width and height of the true target scale,

respectively. In a perfect tracking, the target centre position error is close to 0, and the relative scale is close to 1 [13]. Finally, we also give the time cost of SAMSHIFT.

4.3. Results and Discussion. In this section, we present and discuss the experimental results. We first show the results in the PC environment in Section 4.3.1 and then present the results in the mobile robot in Section 4.3.2. Finally, we discuss the time cost in Section 4.3.3.

4.3.1. Experiments in the PC. As we mentioned, the four algorithms have been implemented in the PC. For each algorithm, it is used to track a black car whose size shrinks over time and is occluded by other cars in the parking area. Figure 7 shows the comparison of the tracking results from the four algorithms.

Figure 7(a) illustrates the tracking results from the traditional Mean-Shift algorithm with the fixed kernel scale. The tracking window drifts around the target area when the size of the target becomes smaller, and the target centre is inaccurate. When the target car is occluded by the white column, the tracking windows drift down largely. After the target car is occluded by the white car, a similar black car parked opposite to the white car is tracked as the target. Figure 7(b) shows the tracking results from the Mean-Shift algorithm with 10% scale adaptation. It can be observed that the tracking results are almost the same as these of the Mean-Shift tracking. Figure 7(c) shows the tracking results by using the spatial color Mean-Shift algorithm. The performance is not good either. When the target is occluded by the white column, the scale of tracking window is much smaller than the true target scale. After the target car is occluded by the white car, the target is lost by this tracker. The tracking results obtained by our SAMSHIFT are shown in Figure 7(d), in which the target is precisely tracked in terms of the scale and the centre position even if the target is occluded by the white column and the white cars.

We manually label the centre position and scales of the moving target for 250 frames. The moving black car, as shown in Figure 7, is subjected to occlusion when passing behind the white column and the white cars.

Figure 8 shows the centre position errors of the target obtained by the four tracking algorithms. As indicated in quantitative comparison in Figure 8, when the black car is occluded by the white car in frame 170, both the Mean Shift and the Mean-Shift with 10% adaption lose the tracking and never recover. The spatial color Mean Shift has a better object representing than the Mean Shift; it loses the tracking at frame 241 and never recovers. SAMSHIFT performs extremely well when the black car is occluded by the white car. In SAMSHIFT, the target candidate generated by the Mean Shift with resizing by the scaling factors loses efficacy, but the target candidate generated by the affine transformation can track the target effectively.

Figure 9 shows the relative scale errors of the target obtained by the four tracking algorithms. As indicated in quantitative comparison in Figure 9, SAMSHIFT adapts to the scale changing very well. In the first 100 frames, the scale of the target is almost fixed, and the target scale obtained by the Mean Shift and the Mean Shift with 10% adaption is close to the true scale before 100 frames. When the target is occluded or the scale of the target is changed, the errors become greater. The target scales obtained by the spatial color Mean Shift are always smaller than the true scale. The target windows become much smaller, and target scales have more errors when the target is subjected to occlusion.

4.3.2. Experiments in the Mobile Robot. In the mobile robot mentioned before, we conduct experiments to test the scale-adaptive tracking performance. The target moves far away from a nearby location. The scale changes from the large size to the small size. The tracking results of the four tracking algorithms are shown in Figure 10.

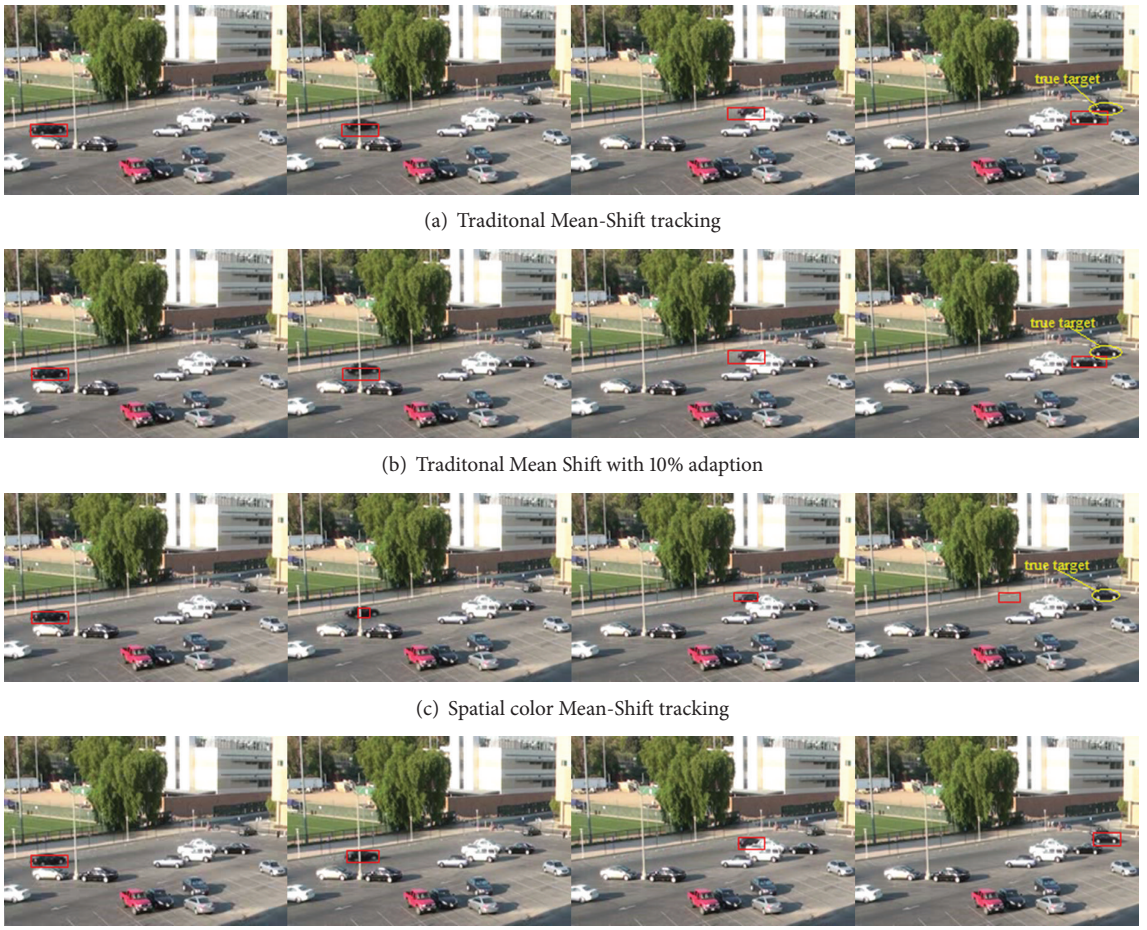
As shown in Figures 10(a) and 10(b), the scales of the tracking windows do not change when we use the traditonal Mean-Shift tracking algorithm and the traditional Mean Shift with 10% adaption. The scales of the tracking windows are incorrect when the spatial color Mean-Shift tracking algorithm is used to track the toy car. Furthermore, the drift becomes larger when the target becomes smaller. Our SAMSHIFT algorithm can adapt to the scale changing very well.

We also manually label the centre position and scales of the moving toy car in Figure 10 for 100 frames. The scale of the moving toy car in Figure 10 changes frequently. Because the video is taken by the ATTC which follows the toy car, the scale of the moving toy car changes from the large size to the small size or from the small size to the large size.

Figure 11 shows the centre position errors of the target toy car obtained by the four tracking algorithms. As indicated in quantitative comparison in Figure 11, SAMSHIFT performs extremely well and has the lowest number of errors. The Mean Shift and the Mean Shift with 10% adaption have more errors. The spatial color Mean Shift has the most errors, which means the tracking windows drift away seriously.

Figure 12 shows the relative scale errors of the target obtained by the four tracking algorithms. As indicated in quantitative comparison in Figure 12, SAMSHIFT adapts to the scale changing very well. The target scales obtained by the other three tracking algorithms have more errors after 20 frames, and there are more and more errors as the target moves. Because the true target scale changes suddenly from the large size to the small size, the cumulative errors of the Mean Shift, the Mean Shift with 10% adaption, and the spatial color Mean Shift increase.

4.3.3. Time Cost of SAMSHIFT. Figure 13 illustrates the time cost of SAMSHIFT. With the target size becoming larger, the time cost increases. However, the curve shows that the time cost is less than 40 ms when the number of the pixels in the target is less than 2×10^4 pixels. That means that SAMSHIFT can be used for real-time tracking applications if there are 25 frames in a second in the video when the target is less than 2×10^4 pixels.



(a) Traditonal Mean-Shift tracking

(b) Traditonal Mean Shift with 10% adaption

(c) Spatial color Mean-Shift tracking

(d) Proposed method

Figure 7: Tracking results' comparison of the four algorithms in PC.

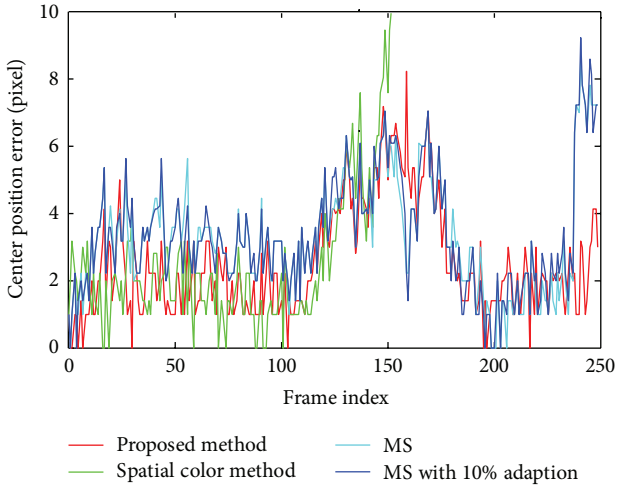


Figure 8: Centre position error comparison.

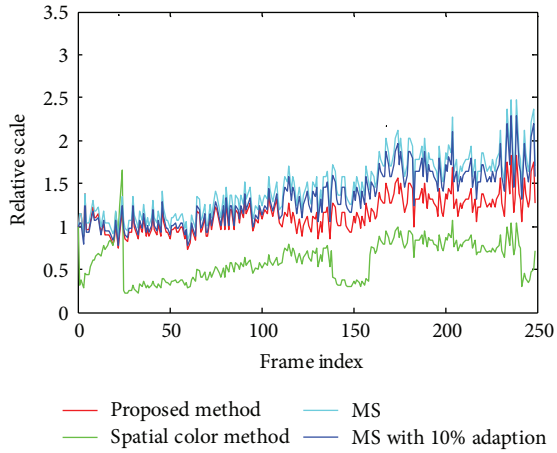
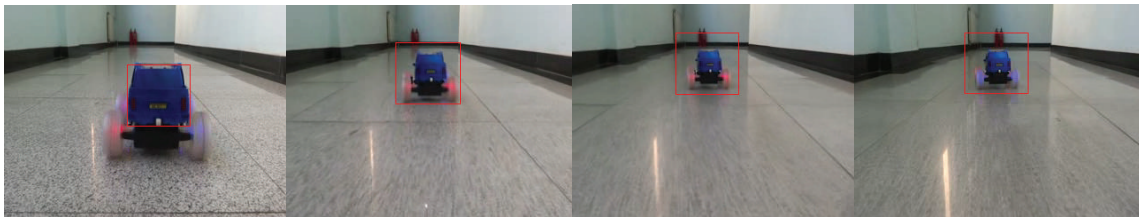


Figure 9: Relative scale comparison.

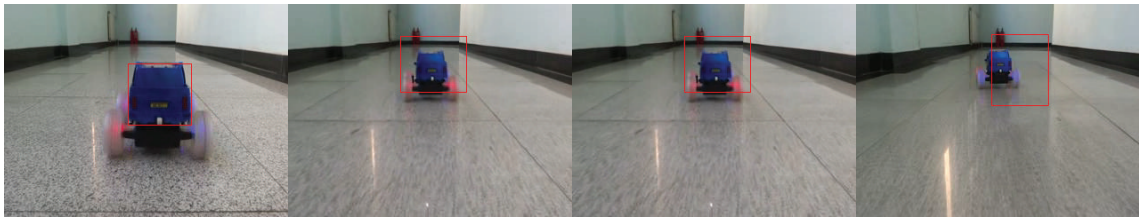
5. Conclusion

We proposed an efficient and effective scale-adaptive tracking algorithm in this paper. In this algorithm, the traditional Mean-Shift tracking algorithm is used to obtain the original target location, and the corner feature matching is used to calculate the affine structure and scaling factor between adjacent frames. In each iteration, three target candidates are

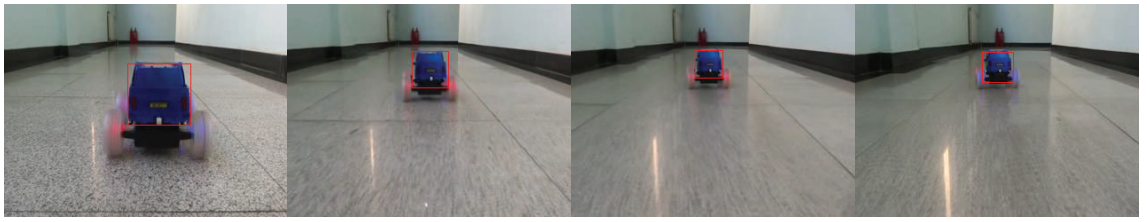
(a) Traditonal Mean-Shift tracking



(b) Traditional Mean Shift with 10% adaption



(c) Spatial color Mean-Shift tracking



(d) Proposed method

Figure 10: Tracking results' comparison of the four algorithms in ATTC.

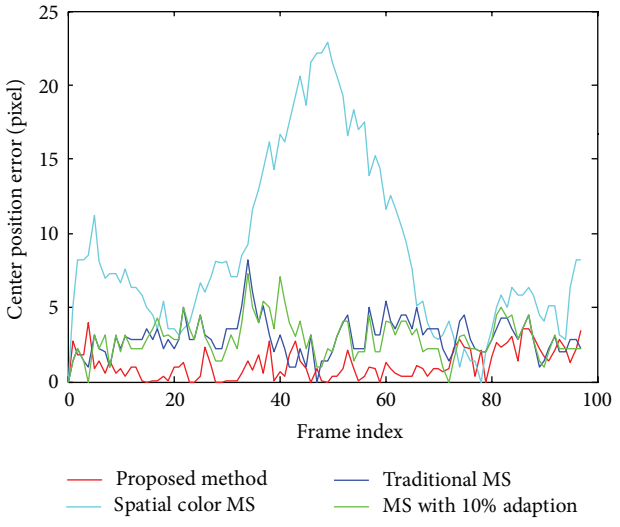


Figure 11: Centre position error comparison.

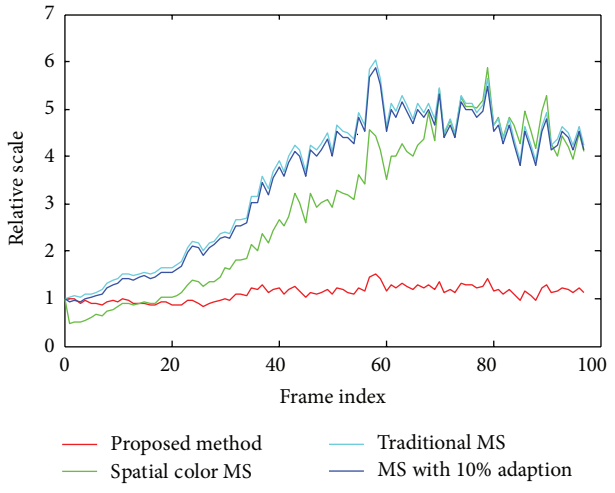


Figure 12: Relative scale comparison.

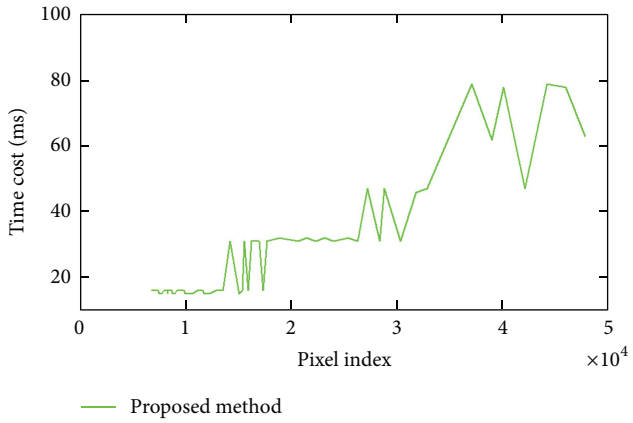


Figure 13: Time cost versus target size.

generated by the affine transformation, the Mean Shift, and the Mean Shift with resizing by the scaling factors, respectively. Then the best candidate is selected. The algorithm has been evaluated in a PC environment and a mobile robot. The experimental results show that our algorithm can effectively adapt to scale changing and is robust to partial occlusion. Furthermore, the tracking speed of our algorithm is fast enough for real-time tracking applications in robot vision.

References

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[2] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Robust mean-shift tracking with corrected background-weighted histogram," *IET Computer Vision*, vol. 6, no. 1, pp. 62–69, 2012.

[3] S. A. Mohammadi, S. Amoozegar, A. Jolfaei, and A. Mirghadri, "Enhanced adaptive bandwidth tracking using mean shift algorithm," in *Proceedings of the IEEE 3rd International Conference on Communication Software and Networks (ICCSN '11)*, pp. 494–498, May 2011.

[4] R. T. Collins, "Mean-shift blob tracking through scale space," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. II-234–II-240, June 2003.

[5] N. Peng, J. Yang, Z. Liu, and F. Zhang, "Automatic selection of kernel-bandwidth for mean-shift object tracking," *Ruan Jian Xue Bao*, vol. 16, no. 9, pp. 1542–1550, 2005.

[6] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," in *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV '98)*, pp. 214–219, 1998.

[7] J. Hu, C. Juan, and J. Wang, "A spatial-color mean-shift object tracking algorithm with scale and orientation estimation," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2165–2173, 2008.

[8] P. F. McLauchlan and J. Malik, "Vision for longitudinal vehicle control," in *Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC '97)*, pp. 918–923, November 1997.

[9] S. Zhu and S. Zhu, "Algorithm of target tracking based on Mean shift with adaptive bandwidth of kernel function," *Opto-Electronic Engineering*, vol. 33, no. 8, pp. 11–16, 2006.

[10] C. Zhao, A. Knight, and I. Reid, "Target tracking using mean-shift and affine structure," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08)*, pp. 1–5, December 2008.

[11] C. Harris and M. Stephens, "A combined corner and edge detector," in *Porceedings of the 4th Alvey Vision Conference*, p. 50, 1988.

[12] J. W. Hwang and H. S. Lee, "Adaptive image interpolation based on local gradient features," *IEEE Signal Processing Letters*, vol. 11, no. 3, pp. 359–362, 2004.

[13] M. Yang, J. Yuan, and Y. Wu, "Spatial selection for attentional visual tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pp. 1–8, June 2007.