

Research Article

Research and Implementation of Sports Entity Simulation Based on Heterogeneous Binocular Vision

Liang Tan ^{1,2} and He Gao³

¹College of Marxism, Jilin Sport University, Changchun, Jilin 130022, China

²College of Physical Education, National Taiwan Sport University, Taoyuan, Taiwan 333325, China

³College of Architecture & Planning, Jilin Jianzhu University, Changchun, Jilin 130118, China

Correspondence should be addressed to Liang Tan; 0333@jlsu.edu.cn

Received 25 February 2022; Revised 18 March 2022; Accepted 28 March 2022; Published 16 May 2022

Academic Editor: Wen Zeng

Copyright © 2022 Liang Tan and He Gao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To study the value of heterogeneous binocular vision in the detection of sports targets, a Zynq-based joint software and hardware design method is proposed, and a mobile object detection system for binocular vision is developed based on it. This paper introduces the relevant technologies and theories used in the system design. First, the basic principles and various modules of the binocular stereo vision process are introduced, including camera tuning, stereo correction, stereo tuning, and telescope stereo vision. Based on the depth information, a method to detect and measure motor targets was developed. Finally, combined with the improved mobile target detection algorithm, the real-time methodology research and algorithm design to determine the target range are completed. The performance of the moving target detection system based on binocular stereo vision and the moving target detection and ranging algorithm are tested and analyzed. We know that the front and rear width of the moving target (the person tested in this paper) is at least 10 cm, while the error value of the ranging algorithm in this paper is within 6 meters, and the average error is less than 10 cm. According to the characteristics of binocular stereo vision, the farther the viewing angle is, the smaller the parallax value of the left and right image pixels will be. Therefore, the insignificant change of depth information will bring some errors to the ranging algorithm. The greater the measurement error is, the lower the percentage deviation of the measurement result is within 2%. It also shows that the target ranging algorithm in this paper can ensure good accuracy within a certain distance. Through this method, we can obtain the accurate data of moving objects in sports competition, so as to improve the training method of athletes and improve the competition results.

1. Introduction

The modern Olympic movement has gone through a glorious history of more than 100 years. While the athletes have won a star-studded medal in the struggle and competition, they have also shown the infinite charm of sports to the world. Through the analysis of the evolution process of Olympic sports performance, the author found that the improvement of almost all sports performance has decreased over time, and even some sports, such as the men's 100 m sports performance, even increased by 1% or 2%. It is very difficult, because human beings rely on their own instinctive power to challenge their own physiological limit, which is close to the limit level. Therefore, only from the perspective

of sports technology, in order to improve or make breakthroughs in sports performance, two changes must be completed in sports technology research methodology, that is, from the traditional one based mainly on human eye observation to one based on high-precision motion capture and analysis. The transformation of the technical measurement method of human movement is from the empirical method based on too much emotion to the human movement analysis method based on the programmed human movement simulation and simulation. As early as 2000, the Institute of Sports Science of the General Administration of Sports of the People's Republic of my country and the Institute of Computer Science of the Chinese Academy of Sciences have successfully developed a series of advanced sports (such as

diving and weightlifting) and quasiadvantage projects (such as gymnastics and trampoline) The “digital three-dimensional human motion simulation system” based on computer virtual simulation technology of intellectual property rights and applying it to the preparations for the Athens Olympic Games not only ensures the absolute advantage of Chinese athletes in diving but also helps the first-generation trampoline athletes in my country. The Olympics won a bronze medal. In the future, China will invest huge human, material, and financial resources in the research and application of sports virtual reality technology based on three-dimensional simulation of sports technology.

Among the many means of human access to information, vision occupies an absolutely important position, and computer vision is to use computers to understand human visual process, in order to replace some difficult tasks with machines [1]. Computer vision is a very complex process. It captures the same scene through two precalibrated cameras and then restores the depth information by calculating the parallax of binocular images on the same spatial point in the same scene and the principle of triangulation. At present, the binocular vision system designed by binocular vision theory has been widely used in aerospace, industrial detection, robotics, automatic driving, and other fields [2]. In recent years, with the deepening of the theoretical research of binocular vision, people’s understanding of the binocular vision system is also deepening. The current research shows that the binocular vision system is not only suitable for indoor scenes but also suitable for outdoor scenes. In the outdoor scene, the depth information of the outdoor scene can be reconstructed through high-resolution binocular camera and appropriate matching algorithm, which greatly enriches the application scene of the binocular vision system. In addition, with the increasing demand for high-performance and low-power computing, there have been many advances in the field of processor architecture. Multicore and multithreaded computing have been widely used in various applications, and heterogeneous multicore processing has gradually emerged in various fields. The multicore processor can greatly improve the computing performance on the premise of limiting the power consumption, and the isomerization of the computing core can make the computing mode of the algorithm not simple serial, but more parallelization, which further shortens the running time of the algorithm. Therefore, the heterogeneous multicore processor has been widely concerned and studied by various industries since its birth. For example, the accelerated processing unit produced by AMD company integrates the traditional CPU periphery with GPU core, which greatly improves the parallel computing ability on the premise of limiting power consumption. The performance of the traditional FPGA processor produced by Zynq company is greatly improved. The number of CPU cores of Xeon series processors produced by Intel company has reached 16, and through the vectorization expansion of CPU instructions, it can also provide powerful parallel computing power; NVIDIA has directly extended GPU to the field of general computing and developed its own CUDA (Compute Unified Device Architecture) programming architecture. Rapid progress is

taking place in the field of high-performance processors and will have a far-reaching impact on traditional computing methods. In this context, the application of heterogeneous multicore processors in the binocular vision system can not only solve the problem of low real-time performance of the algorithm but also effectively control the power consumption. Therefore, binocular stereo vision is also widely used in visual navigation, target detection and tracking, target measurement, target recognition, and 3D scene perception. Figure 1 is a real-time moving target detection system.

2. Literature Review

At present, in competitive sports, the research hotspots of sports technology in various countries focus on two fields: video analysis of sports technology and 3D virtual simulation of sports technology. Among them, the three-dimensional simulation of sports technology is to reproduce the subtle links of the technical movements of elite athletes through computer virtual reality technology. The training intention of the coaches, the organization plan of the manager, and the training process of the athletes achieve the interpretation, analysis, and prediction of the sports system, organization, and evaluation of an experimental technology science. As far as the research and application status of 3D simulation of sports technology in the world is concerned, China is in a relatively leading position, which is mainly due to the national system of my country’s sports competition and management system. To solve this research problem, Fang and others put forward the idea of cost aggregation of bidirectional adaptive weight, which reduces the complexity of cost aggregation, which makes the cost aggregation process of binocular matching the most time-consuming can be completed quickly, and then improves the real-time performance of the matching algorithm [3]. Boulaouche and others proposed the idea of using GPU to accelerate the SGM algorithm based on mutual information in parallel. By using mutual information matching cost and semiglobal optimization method, the edge blur of parallax image caused by low matching accuracy of the local matching algorithm is solved. At the same time, the powerful parallel computing ability of GPU is used, further improving the real-time performance of the semiglobal matching algorithm based on mutual information [4]. GUR and others implemented an adaptive window matching algorithm based on FPGA. The algorithm makes full use of the parallelism and pipelined execution characteristics of FPGA, which not only improves the real-time performance of the algorithm but also reduces the system power consumption. Its disadvantages are long development cycle and poor maintainability of the algorithm [5]. Guo and others evaluated the performance of binocular matching algorithm on various common multicore processors and pointed out that the parallel acceleration effect of GPU on binocular matching algorithm is better than that of DSP and FPGA [6]. Moghaddas and others proposed a sensor method combining radar depth and binocular vision depth, which realizes object segmentation in high-speed environment and has high accuracy and robustness. There are steps of identifying and locating targets [7]. Karray and

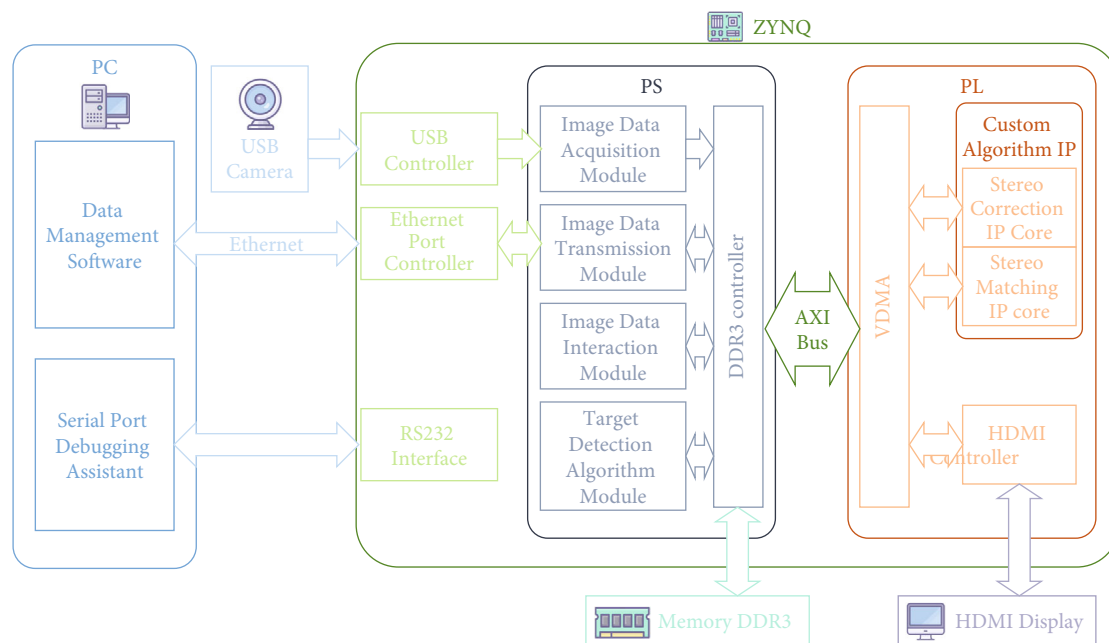


FIGURE 1: Real time moving target detection system.

others proposed LBP (local binary pattern) feature for target detection and face recognition. This method has good effect and simple calculation and is widely used in the field of machine vision [8]. Bieze and others proposed SIFT (scale invariant feature transformation) algorithm on ICCV to extract local features. It uses the found extreme points in space to extract scale, position, and rotation invariants [9]. Sun and others proposed Haar-like feature, which can be used by Lianhart and others to calculate the pixel sum of regional image, which greatly improves the efficiency [10]. Wang and others proposed the hog (direction amplitude histogram) algorithm to extract features. It describes the features through the histogram of statistical image edges and detects pedestrians combined with SVM. This method has strong robustness to illumination changes and improves the recognition accuracy [11]. Jiang and others proposed the surf (accelerated robust feature) feature detection algorithm, which absorbs the advantages of sift and uses the Hessian algorithm to detect key points several times faster than sift [1]. Panda and others proposed DPM (variable component model) target detection algorithm, which improved hog and added multicomponent strategy to describe the target through the relationship between components. The DPM algorithm has become an important part of human behavior and posture classifier [12]. On the current research basis, with the popularity of artificial intelligence all over the world, the science and technology industry also began to develop in the direction of intelligence and digitization. This paper in the aspect of target detection, we should not only find the position of the target in the image but also classify the target objects. This requires the cooperation of target detection and object classification algorithms. In this paper, SSD deep learning neural network combined with mobile net is selected for target detection. Mobile net, as a lightweight clas-

sification algorithm, can quickly and accurately extract the characteristics of target objects. SSD, as a one-step target detection algorithm, can output the position of target objects efficiently and accurately. The combination of the two greatly improves the accuracy of target detection.

3. Method

3.1. Binocular Stereo Vision. The essence of binocular stereo vision is to use two identical cameras to capture two images of the same scene from two perspectives and use the triangulation principle to calculate the position difference between the corresponding pixels of the two images to obtain the parallax map containing the three-dimensional information of the scene [13]. Figure 2 shows the flow of the whole binocular stereo vision technology. Camera data acquisition is the data source of the system, which provides data basis for subsequent processing modules [14]. The application of three-dimensional scene mainly uses the ranging principle of binocular stereo vision and parallax map to complete the application of visual navigation and positioning, object noncontact measurement, and so on. In this paper, it is moving target detection and real-time ranging. Next, it mainly introduces the important part of binocular stereo vision technology.

3.1.1. Camera Calibration. The camera setting is to establish a relationship between the pixels of the image and the actual image location point. The goal is to obtain the camera's internal parameters, external parameters, and distortion parameters, so as to lay the foundation for the use of the stereoscopic editing module and the following 3D scenarios. The imaging principle of the aperture camera determines the transition process between the four coordinate systems.

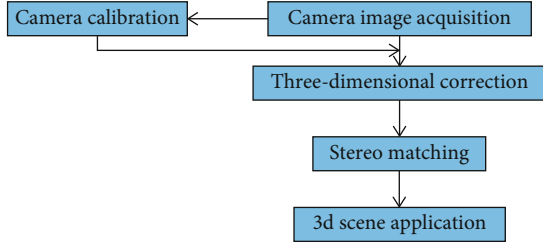


FIGURE 2: General flow of binocular stereographic techniques.

Assuming that the coordinate of a point in the world coordinate system is $[x_w, y_w, z_w]^T$ and the imaging point in the image pixel coordinate system is $[u, v]^T$, the conversion relationship of the point from the world coordinate system to the image pixel coordinate system is

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} RT \\ 01 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}. \quad (1)$$

Here, R is an orthogonal rotation matrix, and T is a translation matrix. They are used to determine the positional relationship between the global and camera frames and the relative position of the telescope camera. The parameters required to define the two matrices are called external parameters. u_0 and v_0 represent the position coordinate of the center of the virtual imaging plane in the image pixel coordinate system. f_x represents the projection of the focal length in the X direction in the image physical coordinate system, and f_y represents the projection in the Y direction. $f_x, f_y, u_0,$ and v_0 are the internal parameters of the camera. Internal parameters indicate how to convert 3D camera coordinates into 2D image coordinates inside the camera. Therefore, internal parameters are only related to the camera itself. Another internal parameter of the camera is distortion parameter. Distortion is the distortion and change caused by the camera in order to collect more image data under limited viewing angle. It generally includes radial distortion and tangential distortion. A total of five parameters need to be determined.

Among the calibration methods, the calibration method is simple and mature. In the process of calibration, only one chess and card grid panel needs to be used as the calibration board. By fixing the camera, let the calibration plate move at different positions and angles to collect the image of the camera [15]. Different equations are established according to the key points on the calibration board, and then the values of relevant parameters are obtained by closed-form solution and maximum likelihood estimation method.

3.1.2. Stereo Correction. The first step of stereo correction is image correction, which uses the camera distortion parameters to remove the distortion of the image. In stereo correc-

tion, the most mature method is Bouguet's stereo correction algorithm. The principle of this algorithm is the epipolar correction method. The principle of this method is briefly introduced below.

Due to the epipolar geometry of the corresponding pixels of the binocular image, under the epipolar constraint, for the feature points on the imaging plane, the matching points on the other imaging plane must be on the corresponding epipolar line. The function of binocular stereo correction is to make strict line correspondence between the two corrected images by using the geometric relationship of the opposite level and keep the epipolar lines of the two images on the same horizontal line, so that a corresponding pair of pixel points can be found on the same line of the two images [16]. When calculating the image parallax to find the corresponding point of the pixel point, it only needs to carry out linear search in this line, to speed up the calculation speed, and to reduce the false matching rate.

3.1.3. Stereo Matching. Its purpose is to match the corresponding pixel points in the binocular image after stereo correction, calculate the difference of u coordinates of these corresponding points in the image pixel coordinate system in the left and right images, obtain the difference of corresponding points, and finally combine the difference of all pixel points to form a parallax map.

Stereo matching algorithms are complex and diverse. The research on stereo matching algorithms is also a hot research direction in binocular stereo vision technology. In these three categories, researchers have proposed a variety of stereo matching methods. The most commonly used is region-based matching. Its principle is to establish a region block window centered on the pixel to be matched in the reference image and then find a pixel in another image, so that the region blocks window of the same size created with it as the center and the region block window in the reference image meet the similarity under certain threshold conditions.

3.1.4. Principle of Binocular Stereo Vision Ranging. According to the principle of similar triangle, the depth information Z can be deduced:

$$\frac{|x_l - x_r'|}{B} = \frac{f}{Z} \Rightarrow Z = \frac{fB}{d}, \quad (2)$$

where B represents the distance between the optical centers of binocular cameras, also known as the camera baseline distance. The parallax value D of the target point P can be obtained through the stereo matching algorithm, and the depth information of the point P , that is, the distance value, can be obtained by bringing its value into equation (2).

3.2. Software and Hardware Collaborative Design. Zynq is a chip combining software and hardware. Software and hardware collaborative design aims at the system requirements. In the design process, software and hardware interact to achieve the purpose of efficient work of the system. In the early development of software and hardware system,

hardware and software development were divided into two independent parts. Generally, the hardware is designed first, and then the software is designed on the hardware platform. Due to the lack of clear understanding of software and hardware architecture and implementation mechanism in the design process, the design results are often blind. As a result, the whole system design needs to rely on the designer's development experience, and the design time and cost are greatly improved. Moreover, in the process of repeated modification, it often deviates from the original design requirements in some aspects. Compared with the disadvantages of the traditional independent design of software and hardware architecture, the software and hardware collaborative design method excavates the correlation between the system software and hardware as much as possible by comprehensively analyzing the functions [17]. Divide the software and hardware reasonably, then complete the software and hardware data interaction in a correct way, and finally generate the hardware and software architecture of the system through hardware comprehensive simulation and software compilation and testing.

In the design process, the functional requirements of the system are analyzed from the description of the target system. Considering the existing resources, algorithm complexity, and cost and time of system development, the software and hardware of the system are divided reasonably, make clear which module is implemented by hardware and which module is implemented by software, and design the software and hardware interaction interface. Secondly, the software and hardware of the system are developed synchronously, including the synthesis and simulation of hardware modules, the compilation and testing of software modules, and the interactive design of software and hardware interfaces. Finally, the integrated hardware is integrated with the compiled software, and the simulation test of the system is carried out to detect whether the system meets the performance requirements. Otherwise, it is necessary to redivide the software and hardware. With the increasing development and maturity of programmable devices, excellent embedded systems are inseparable from the architecture method of software and hardware collaborative design.

3.3. Traditional Sports Object Detection Algorithm. By analyzing the sequence of video or continuous image frames, determine whether each frame has a forward target, that is, moving the target; then decompose the corresponding front-end target attribute, i.e., the detection of the moving target is completed. Based on the presence of relative motion between the detected phenomenon and the camera, the motion target detection algorithm can be divided into static background detection and motion background detection [18].

3.3.1. Frame Difference Method. The frame difference method uses the principle of strong correlation between adjacent frames of image sequence to separate moving targets through the differences between image frame sequences. Its essence is to calculate the difference of pixel values between two or more adjacent frames in image sequence

and set the corresponding threshold to divide the foreground region and background region.

Assuming that the k -th frame image in the continuous image sequence is represented as $f_k(x, y)$, the K - T frame image is represented as $f_{k-t}(x, y)$, T represents the interval of the differential frame, and the difference $D_k(x, y)$ of the two frame images can be described by formula (3):

$$D_k(X, y) = |f_k(X, y) - f_{k-t}(x, y)|. \quad (3)$$

Then, the $D_k(x, y)$ is binarized, and the result is shown in equation (3). By analyzing the $M_k(x, y)$ value to distinguish the front scenic spot,

$$M_k(x, y) = 1, D_k(x, y) > T, \quad (4)$$

$$M_k(x, y) = 0, D_k(x, y) \leq T. \quad (5)$$

The moving target can be extracted from the background by binarization of all pixels of the target image. The flow chart is used to describe the frame difference method, as shown in Figure 3.

In Figure 3 that after the image difference is calculated through equations (3) and (4) and binarized, some methods in morphology (such as filtering) need to be used to attenuate the noise in the binarized image. Because the moving target obtained by frame difference method may have some holes, it also needs connectivity processing, which finally determines the moving target.

The algorithm of frame difference method has low complexity, can better adapt to the background environment with slow change, and has high real-time performance. However, the detection accuracy of the algorithm is not high. When the color of the target is like the background, it cannot effectively detect the moving target; moreover, the algorithm is prone to false detection. For example, when checking fast-moving targets, it is easy to detect multiple moving targets or treat the background as moving targets.

3.3.2. Background Subtraction Method. Background subtraction is the most used basic method in moving target detection. Its essence is to establish a background model to represent the detection scene; then, the pixels of each subsequent frame are subtracted from the corresponding pixels in the background, and the ones with large changes in the operation results are regarded as the front scenic spots, and vice versa. The core of the background subtraction method is the establishment of the background model, which should reflect the realistic changes of the scene as much as possible. Because the real scene cannot be static, the factors such as illumination, water wave, and the change of leaves all increase the difficulty of background modeling, and the background model needs to be updated constantly. There are many modeling methods for different application environments. Figure 4 shows the working flow chart of background subtraction method for primary target detection.

By analyzing the flow of Figure 4, the algorithm first needs to obtain a frame of image in the image sequence and attenuate the image noise through simple preprocessing

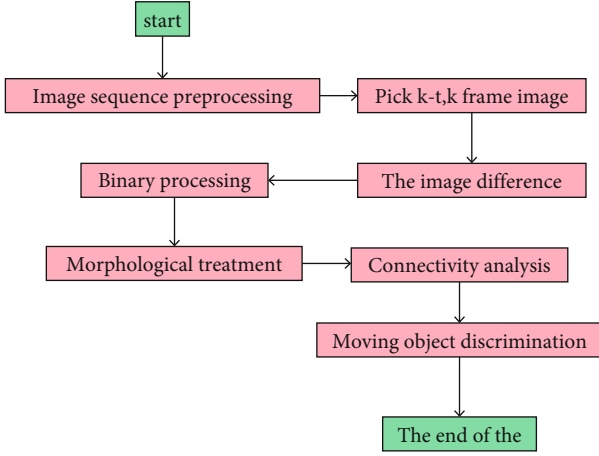


FIGURE 3: Workflow of the frame difference method.

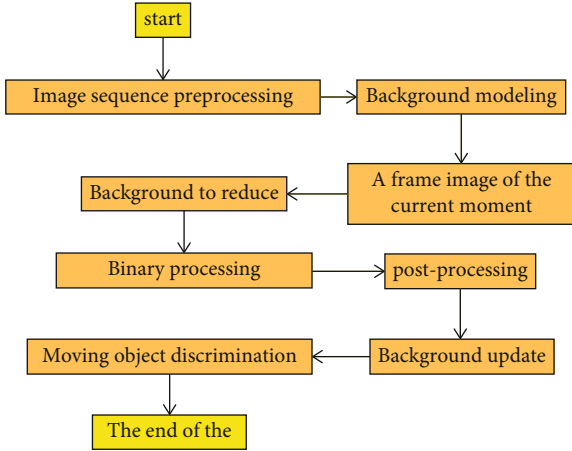


FIGURE 4: Workflow of the background subtraction method.

and then carry out background modeling according to a certain background modeling principle. If a frame image at time t is selected as $F(x, y, t)$, assuming that the image after background modeling at this time is $B(x, y, t)$, the background of the image is subtracted to obtain the difference between the image and the background. By comparing with the threshold T selected in a specific environment scene, the binarized image $I(x, y, t)$ is obtained. The expression is as follows:

$$\begin{aligned} I(x, y, t) &= 1 |F(x, y, t) - B(x, y, t)| > T, \\ I(x, y, t) &= 0 |F(x, y, t) - B(x, y, t)| \leq T. \end{aligned} \quad (6)$$

The binary image has highlighted the foreground target. In order to eliminate noise, it is also necessary to carry out morphological filtering such as frame difference method and target connectivity analysis. Because the background model is not one layer invariant, corresponding algorithms need to be adopted to update the background model and finally determine the moving target.

It has little computation and is easy to implement and performs well in obtaining the foreground area when the scene changes slightly; however, the algorithm is sensitive to illumination changes and vulnerable to background fluctuations. The effect of this method also depends on the establishment and updating of the background model. At present, the commonly used background updating methods include Gaussian mixture modeling and vibe model updating. Due to the large amount of calculation and high complexity of Gaussian mixture modeling, it is not suitable for the moving target detection system with high real-time requirements; so, it will not be described in detail here.

3.3.3. Vibe Algorithm. The Vibe algorithm is a pixel-level video background modeling algorithm. The algorithm stores a sample set of all pixels in an image, and each sample set is the pixel value before the pixel and the pixel value of the neighbors; then, compare the new range with the value in the sample set to determine whether the point is a background point and update the sample value of the pixel and its field points with a certain probability. The algorithm has no limited requirements for the detection environment. It estimates the background model through random sampling to simulate the random volatility of the real environment. By adjusting the time secondary sampling factor, a few new sample values are used to replace the changes of all sample values, considering the detection accuracy and computational complexity.

For any pixel $v(x, y)$ in the selected image, the algorithm randomly selects the pixel values $v_i(x, y)$ of N neighborhood points as sample values and stores them in the background sample set $M(x, y)$ of the pixel. The result is shown in formula (7). Finally, the sample values of all pixels in the image are collected to initialize the background model.

$$M(x, y) = \{v_1(x, y), v_2(x, y), \dots, v_N(x, y)\}. \quad (7)$$

Positive image detection is as follows: starting from the second frame of the video image sequence, go through each pixel of the new frame image in $v(x, y)$ to modify the pixel sample set:

$$D(x, y) = (d_1(x, y), d_2(x, y), \dots, d_N(x, y)), \quad (8)$$

where $D(x, y)$ represents the set of differences between the value of the current pixel $v(x, y)$ and all sample values in the pixel sample set $M(x, y)$:

$$d_i(x, y) = v(x, y) - v_i(x, y). \quad (9)$$

Before foreground detection, the threshold r is to judge whether the pixel is close to the historical sample value, and the threshold T is to judge whether it is the front scenic spot need to be set in advance; then, traverse all elements $d_i(x, y)$ in $D(x, y)$ and compare its size with the threshold R ; count the number of $d_i(x, y)$ as N_f . If $N_f > T$, this point is the front scenic spot. According to the experimental test results, the three main parameters in the detection process

are generally set as follows: the number of sample sets $N = 20$, the threshold $R = 20$, and the threshold $T = 2$.

Background model updating is as follows: while detecting the foreground, the background model needs to be updated constantly so that the model can adapt to the changing scene. The update method determines that the update strategy of the vibe algorithm includes three attributes: memoryless update strategy, time sampling update strategy, and spatial domain update strategy. First, whether the sample values are replaced in the sample set of the background model is independent of time. Secondly, the algorithm reduces the update frequency of the sample value in the background and its neighbor sample value through the time secondary sampling mechanism. Third, for the pixel to be updated, a pixel will be randomly selected in its neighborhood to update the selected sample pixel value with the newly selected pixel [19].

According to the understanding of the basic principle of the vibe algorithm, the algorithm process is simple, has strong randomness in time and space, and can effectively adapt to environmental changes. The vibe algorithm is more stable than the frame difference method and background subtraction method and is not easy to be affected by illumination and scene mutation; at the same time, the algorithm has low complexity and small amount of calculation, which is easy to implement on embedded devices [20]. Through the test of the data set, it is found that if the scenes in frame are all background, the detection effect will perform well, but if there are moving objects in the first frame, the phenomenon of “ghost” will occur, or the sudden movement of stationary targets in the process of moving target detection will also produce the phenomenon of “ghost.” Secondly, the detection of moving target by the vibe algorithm is usually incomplete, resulting in that the moving target is not a connected region in the detection effect. These two problems will have an adverse impact on the target ranging module of the moving target detection system: “ghost” phenomenon will lead to multiple target areas, resulting in false detection and blank ranging. If the target detection is incomplete, there will be multiple subregions and multiple inaccurate distance information. In view of these two shortcomings of the vibe algorithm, this paper improves it according to the depth information obtained by binocular stereo vision technology [21].

3.4. Improved Vibe Algorithm Based on Depth Information

3.4.1. Correction of “Ghost” Phenomenon. If the frame image used to initialize the background model happens to have a moving target, it will be regarded as the background. Similarly, when the target is forbidden to move suddenly during the detection process, such a “ghost” phenomenon will also appear. Considering that the depth information of pixels in this area will increase suddenly after the target moves, combined with this feature, when the algorithm initializes the background model, the depth information of each pixel will also be stored in the sample set. In a new frame of image detection, the difference value of depth information will be increased, and the result information is stored in the matrix mat D :

$$\begin{aligned} \text{Mat}D(x, y) &= 1 | \text{depth}_s(x, y) - \text{depth}(x, y) | > D, \\ \text{Mat}D(x, y) &= 0 | \text{depth}_s(x, y) - \text{depth}(x, y) | \leq D, \end{aligned} \quad (10)$$

where depth represents the $\text{depth}_s(x, y)$ information in the pixel sample set, depth_s represents the $\text{depth}_s(x, y)$ information of the current pixel, and D represents the differential threshold of the depth information. It is necessary to correct the point as a background point and reinitialize the background model.

3.4.2. Correction of Incomplete Target Detection. This phenomenon is usually caused by the complexity and variability of the scene or moving target. There is also some noise interference, which makes isolated noise points and connected noise areas appear in the image data. Considering that the depth information of pixels in the moving target area is similar, there will not be too much mutation.

First, you need to define eight field value sets $s(x, y)$ of any pixel (x, y) :

$$\begin{aligned} S(x, y) &= \{ (x-1, y-1), (x, y-1), (x+1, y-1), (x-1, y-1) \\ &\quad \cdot (x+1, y), (x+1, y-1), (x, y+1), (x+1, y+1) \}. \end{aligned} \quad (11)$$

Then, the depth information of the pixel (x, y) and the depth information of the field value are differentiated:

$$| \text{depth}(x, y) - \text{depth}_i(x, y) | \leq F, \quad (12)$$

where $\text{depth}(x, y)$ represents the depth information of the pixel, $\text{depth}_i(x, y)$ represents the depth information of the neighborhood of the pixel, and f represents the difference threshold of the depth information between the two. If the current pixel point is the front scenic spot and its neighborhood point satisfies formula (12), its neighborhood point is also corrected as the front scenic spot.

3.5. Target Ranging Algorithm. We know that moving object detection with distance information can effectively improve the safety of pedestrians and vehicles in road traffic. However, the common target ranging is usually to detect the feature points of moving targets and express the target distance through the single depth information of feature points. This method is simple and easy, but the selection of feature points often cannot fully include all the information of the target. The target is a moving region, and several key feature points cannot be used to describe the distance information of the region. To solve this problem, this paper analyzes the depth information of the moving target area, classifies, discards all the depth values in the area, and finally calculates the mean value of the selected value to represent the real-time distance of the moving target.

Through the improved vibe algorithm described in the previous section, the region of the moving target can be accurately obtained, and then the moving target can be framed with a rectangle by using the simple method in openCV, considering that the depth information of all pixels in the moving target contour should have certain similarity,

even the same value. After removing the moving target in the rectangular area, the background is the background. The depth information of the background points has similar similarity, and their values must be much larger than all pixels in the moving target contour; moreover, the number of depth values of foreground targets in the detection area should be much greater than the number of depth values in the background. According to these characteristics, this paper classifies and discards the depth information in the rectangular box.

Suppose that the starting pixel of the rectangular box is (u_0, v_0) and the ending pixel is (u_i, v_i) , define the maximum distance as Max, divide the distance value from 0 to Max into equal intervals, totaling n segments, and the interval of each segment is d ; Then, the depth value of each interval $(k, k + d)$ is statistically summed:

$$S_i = \sum_{(x,y)=(u_0,v_0)}^{(x,y)=(u_i,v_i)} \text{depth}(x, y) \text{depth}(x, y) \in (k, k + d), \quad (13)$$

where S_i represents the sum of the depth values of each interval, $\text{depth}(x, y)$ represents the depth value of pixels (x, y) , $k = 0, d, 2d \dots \text{MAX}-d$. At the same time, count the number of depth values in the interval $(k, k + d)$:

$$C_i = \sum_{(x,y)=(u_0,v_0)}^{(x,y)=(u_i,v_i)} 1 \text{depth}(x, y) \in (k, k + d), \quad (14)$$

where C_i represents the sum of the number of depth values of each interval. Then, put the S_i and C_i of each interval into the set, respectively:

$$\begin{aligned} \text{ConA} &= \{S_1, S_2, \dots, S_N\}, \\ \text{ConB} &= \{C_1, C_2, \dots, C_N\}. \end{aligned} \quad (15)$$

After the completion of statistics, calculate the percentage of each element in the set con B in the sum of elements in the whole set. Select the element C_i with the highest percentage value and find the corresponding S_i in the set con A according to the index value I . Finally, the mean S_i/C_i is obtained. The distance of the moving target obtained by this algorithm can effectively avoid the deviation caused by the method of estimating the distance by local feature points and has a certain accuracy in the test system.

4. Results and Analysis

Due to the bandwidth limitation of USB2.0 interface of development board, the system cannot meet the real-time data acquisition of camera. In order to meet the test requirements, the system test adopts USB3.0 interface of PC, and the collected data is quickly transferred to the installed hardware card through PCIe interface, which can avoid data delay and detection target and improve the accuracy of the test system algorithm. Ethernet port transmits data to PC through UDP transmission thread for later storage and

TABLE 1: System hardware resource consumption.

Resource	Utilization	Available
LUT	182789	218601
FF	171530	437201
BRAM	485	544
DSP	267	901

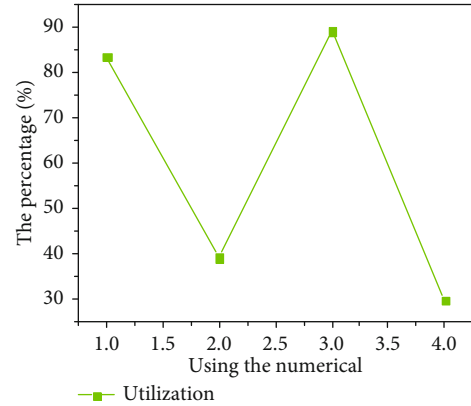


FIGURE 5: System referral resource utilization.

TABLE 2: Time of complex algorithms.

Platform	Stereo correction	Stereo matching	Total
CPU	0.230 s	29.752 s	29.982
Zynq	0.00472 + 0.00351		0.00823 s

application; PCIe interface is used for high-speed transmission of data collected by camera. Based on this experimental platform, the following first introduces the environment construction and performance test of the system, and then introduces the result test and data analysis of the moving target detection and ranging algorithm of the system.

4.1. System Construction and Performance Test. First, the mobile target detection system implements the PL logic design section through the Xilinx Vivado development toolkit, which can process and encapsulate the IP kernel. Using this tool, we can complete the design of the specific IP kernel design in the system, the hardware path between the IP cores, and verify that the design is optimal and correct with the help of hardware simulation. As shown in Table 1 and Figure 5, the main resources spent on the system hardware are as follows.

As can be seen from Table 1 and Figure 5, the main resource consumption of the development board is within the available range and can meet the system operation requirements. The PL part realizes the stereo correction and stereo matching algorithm through the parallel acceleration of FPGA to improve the processing speed of the system.

TABLE 3: Comparison results of system processing frame rate.

Design architecture	Device	Resolving power	Hardware resource	Frame rate
Design of embedded real-time	XC7Z045	640*480	182933 LUTS, 143223 FFS, 100 DSPs, 138 BRAMs	101
This paper	XC7Z045	640*480	182796 LUTS, 171526 FFS, 268 DSPs, 485 BRAMs	121.42

TABLE 4: Frame rate of image processing with different resolutions.

Resolving power	640*480	1280*720	1280*960	1920*1080
Frame rate	121.426	66.725	54.959	36.560

To obtain a speed comparison between the hardware and software design architecture (fpga100 MHz, arm667 MHz) and simple CPU (2.50 GHz), the processing times were checked and collected using a telescope image test at 640 * 480 resolution. Two complex algorithms are multiple stereo corrections and stereo tunings. The PS part calls the system clock according to the program instructions and calculates the software development time through the jet lag. Table 2 shows the average time to develop the test package. It can be seen that a simple CPU takes almost half a minute to develop the two algorithms, while the Zynq system develops the two algorithms in parallel according to the collaborative design method of the architecture. Logical control of the PS part takes only 0.00473 seconds, while the algorithm development of the PL part takes only 0.00350 seconds.

This paper compares the system architecture with the embedded real-time design stereo vision system architecture only for the processing time of complex algorithms. The comparison results are shown in Table 3. Although the embedded real-time design stereo vision system architecture has low hardware resource consumption, however, the image frame rate is slightly lower than the architecture method proposed in this paper. Moreover, the system in this paper also reflects the human-computer interaction performance, while the embedded real-time design stereo vision system does not reflect any human-computer interaction performance.

The development board does not support the rapid transmission of high-resolution camera and cannot test the best effect of the overall system. Therefore, the acquisition end is changed to store atlas with different resolutions in the startup card. The system carries out actual test by reading these atlantes and controls the interactive work such as system startup and resolution change through buttons on the GUI. In PS coding, the method of calling system time through program instructions calculates the time difference before receiving data from PL. The test results are shown in Table 4 and Figure 6:

It can be seen from the table and figure that when processing images with a resolution of 640 * 480, the frame rate of the system can reach 121.44 frames/s. The real-time effect can be achieved by displaying the processing results through HDMI display and GUI, if the system is applied to vehicle safe driving.

4.2. *Application Algorithm Test and Result Analysis.* Disparity map information technology can meet many application

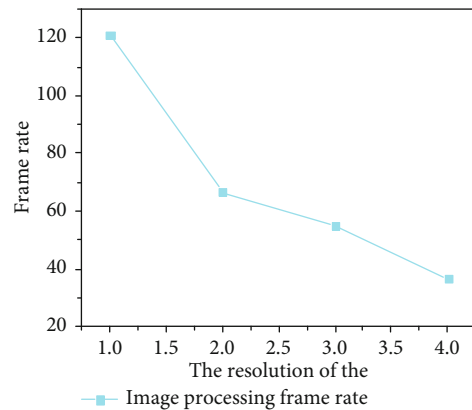


FIGURE 6: Frame rate diagram of image processing with different resolutions.

requirements in 3D scene perception. In order to detect the sports entity target, this paper completes the application of moving target detection. The ZC706 hardware development board selected for the system adopts a low-level manual system with low processing ability and low real-time of the application algorithm. Therefore, the effect of the system application algorithm is tested on the computer [22].

In this paper, the video with image resolution of 640 * 480 is sampled and tested in indoor lighting environment and outdoor brightness uniform environment, respectively. First, build the system test platform correctly. Then, with the help of the ruler, let the target walk parallel to the camera angle at the fixed actual distance point and collect and process the video to ensure that the camera captures the moving target at the same distance. Finally, the system test distance of each frame image is recorded.

In order to minimize the measurement error, the measurement values of all frame images from frame 100 to frame 300 in the measurement video of each distance point are counted in this paper. Then, compare these measured values and select the target value with the largest number of measured distance values as the system measurement value of the moving target. Based on this statistical method, the indoor environment and outdoor environment are measured and counted, respectively. The statistical results are shown in Figures 7–14:

Because the camera height setting of the system platform is different when it is built indoors and outdoors, this paper

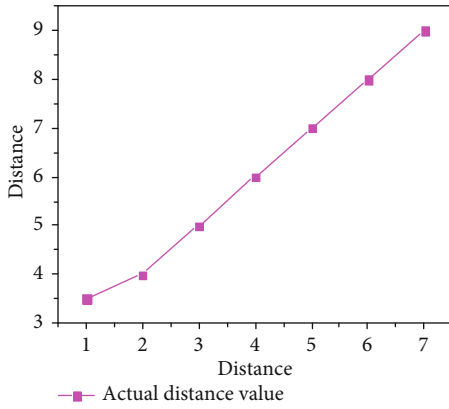


FIGURE 7: Actual value of indoor moving target ranging.

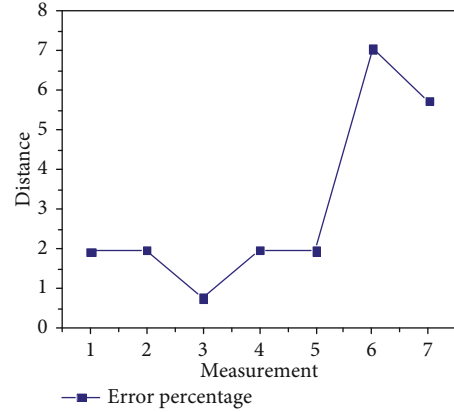


FIGURE 10: Indoor moving target ranging error percentage.

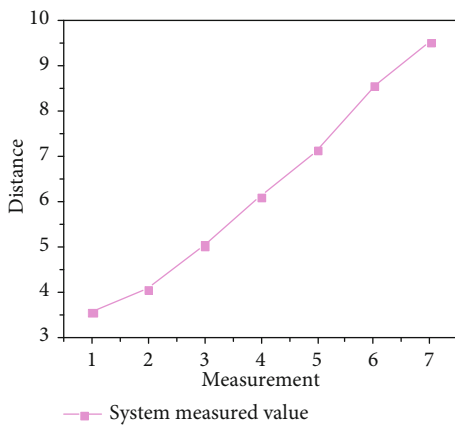


FIGURE 8: Measured values of the indoor moving target ranging system.

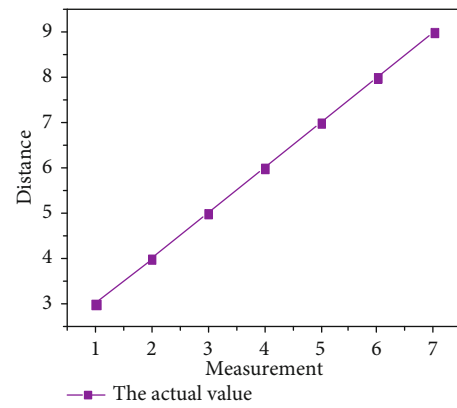


FIGURE 11: Actual value of outdoor moving target ranging.

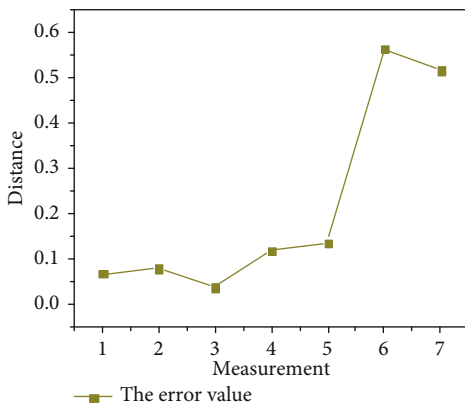


FIGURE 9: Indoor moving target ranging error value.

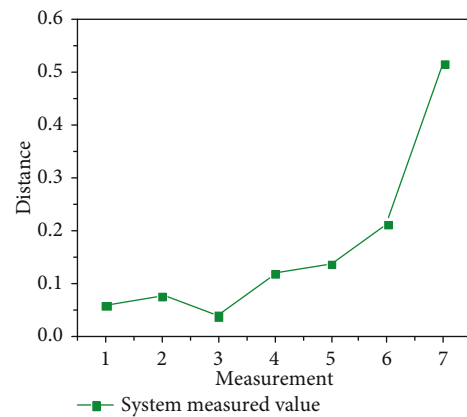


FIGURE 12: Measured values of the outdoor moving target ranging system.

takes the ability to see the whole contour of the moving target as the standard for distance measurement; so, the starting point of indoor ranging is 3.5 m, and the starting point of outdoor ranging is 3 M. Moreover, the moving target has become small and blurred beyond 9 meters; so, the measurement ends at 9 meters. From the comparison of Figures 7–14, in this article, you can see that the indoor

and outdoor performance is slightly different for the target algorithms. We know that the moving object (the person tested in this document) is at least 10 cm wide, but the algorithm error specified in this document is within 6 m, and the average error is less than 10. From the stereo visual characteristics of the binocular, the farther the perspective, the smaller the pixel parallax of the left and right images.

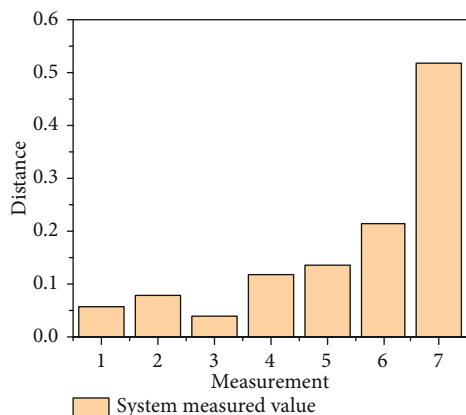


FIGURE 13: Ranging error value of outdoor moving target.

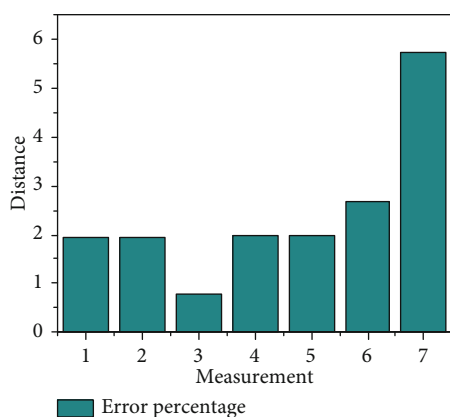


FIGURE 14: Percentage of ranging error of outdoor moving target.

Therefore, small changes in depth information can all lead to some algorithmic errors. It can be seen that the higher the measurement error, the smaller the measurement deviation is less than 2%; it also shows that the target distance algorithm can provide good accuracy within a certain distance.

5. Conclusion

This paper presents the research and implementations of various binocular-based moving object modeling. To detect objects of moving objects, we complete a program to detect moving targets and determine real-time distances based on the traditional vibration algorithm and scaling stereo visual information. The ZC706 hardware development board selected for the system implementation adopts a low-level manual system, with low processing ability and low real-time performance of the application algorithm. Therefore, the effect of the system application algorithm is tested on the computer. This paper presents a software and hardware codesign method for the moving target detection system. Combine the respective advantages of software and hardware to divide the software and hardware of the system module and then design efficient and accurate data control and transmission between software and hardware; finally,

combined with the specific algorithm and logic control of binocular stereo vision technology and moving target detection technology, the software and hardware of the system are realized. In the test and analysis of the real-time performance of the system, the processing speed of the system decreases linearly with the increase of image resolution. In the future development of science and technology, people's requirements for image quality are gradually improved, and there is a great demand for high-resolution images. Therefore, the system can be further optimized in the architecture design and binocular stereo vision algorithm, such as increasing the bandwidth of data transmission interface under high resolution in the architecture, to reasonably reduce the complexity of the algorithm when ensuring the accuracy of the algorithm. Through this method, we can obtain the accurate data of athletes in sports competition. Through data analysis, we can improve the athlete training method, so as to improve the competition performance. The experiment shows that this method can be effectively applied to sports training.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no competing interests.

Acknowledgments

This work is supported by the National Social Fund of China, A Study of Entertainment Culture in Tang Dynasty (18CTY016).

References

- [1] S. Jiang, P. Lin, Y. Chen, and C. Tian, "Research on heterogeneous fiber grating spectrum separation based on transfer matrix and adaptive algorithm," *Optical Engineering*, vol. 58, no. 9, p. 1, 2019.
- [2] W. Song, M. Xu, and Y. Dolma, "Design and implementation of beach sports big data analysis system based on computer technology," *Journal of Coastal Research*, vol. 94, no. sp1, p. 327, 2019.
- [3] Z. Fang, H. Gao, and Z. Ma, "Prediction and evaluation of groundwater environmental impact based on numerical simulation model," *Journal of Coastal Research*, vol. 115, no. sp1, p. 286, 2020.
- [4] T. Boulaouche, D. E. Kherroub, K. Khimeche, and M. Belbachir, "Green strategy for the synthesis of polyurethane by a heterogeneous catalyst based on activated clay," *Research on Chemical Intermediates*, vol. 45, no. 6, pp. 3585–3600, 2019.
- [5] S. Gur, G. N. Frantziskonis, S. Pannala, and C. S. Daw, "Application of wavelet-based methods for accelerating multi-time-scale simulation of bistable heterogeneous catalysis," *Industrial & Engineering Chemistry Research*, vol. 56, no. 9, pp. 2393–2406, 2017.

- [6] Y. Guo, Z. Lu, Q. Wang et al., "Detecting density dependence from spatial patterns in a heterogeneous subtropical forest of Central China," *Canadian Journal of Forest Research*, vol. 45, no. 6, pp. 710–720, 2015.
- [7] M. Moghaddas and A. Davoodnia, "Atom-economy click synthesis of tetrahydrobenzo [b] pyrans using carbon-based solid acid as a novel, highly efficient and reusable heterogeneous catalyst," *Research on Chemical Intermediates*, vol. 41, no. 7, pp. 4373–4386, 2015.
- [8] S. Karray, "Modeling brand advertising with heterogeneous consumer response: channel implications," *Annals of Operations Research*, vol. 233, no. 1, pp. 181–199, 2015.
- [9] T. M. Bieze, A. Kruszewski, B. Carrez, and C. Duriez, "Design, implementation, and control of a deformable manipulator robot based on a compliant spine," *The International Journal of Robotics Research*, vol. 39, no. 14, pp. 1604–1619, 2020.
- [10] Y. Sun and X. Zhao, "Research and implementation of license plate recognition based on android platform," *MATEC Web of Conferences*, vol. 309, no. 15, article 03034, 2020.
- [11] M. Wang, H. Zhang, Q. Hu, D. Liu, and H. Lammer, "** Research and implementation of a non-supporting 3D printing method based on 5-axis dynamic slice algorithm**," *Robotics and Computer-Integrated Manufacturing*, vol. 57, pp. 496–505, 2019.
- [12] A. R. Panda, D. Mishra, and H. K. Ratha, "Fpga implementation of software defined radio-based flight termination system," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 1, pp. 74–82, 2015.
- [13] H. Li, "Research and implementation of art animation system based on video stream," *Revista de la Facultad de Ingenieria*, vol. 32, no. 6, pp. 153–161, 2017.
- [14] L. Li, "Research and implementation of fast image processing algorithm based on fpg," *Revista de la Facultad de Ingenieria*, vol. 32, no. 3, pp. 749–756, 2017.
- [15] Y. Peng, X. Liu, G. Yu, W. Wang, and J. Wang, "Research and implementation of localization algorithm based on zigbee technology," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 4, pp. 9–15, 2017.
- [16] W. Fang, "Research and implementation of key technologies of digital forensics based on cloud computing," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 6, pp. 572–579, 2017.
- [17] Y. Liu, "Research on heterogeneous data fusion algorithm based on iot," *Revista de la Facultad de Ingenieria*, vol. 32, no. 4, pp. 549–556, 2017.
- [18] D. Ou, L. Chen, L. Shi, X. He, and D. Dong, "Research on speed-guidance control model and implementation method for tramcars based on signal priority at intersections," *Tiedao Xuebao/Journal of the China Railway Society*, vol. 39, no. 6, pp. 80–86, 2017.
- [19] Y. C. Song, J. Xiong, L. I. Xian-Wang, and X. H. Chen, "Integration for heterogeneous manufacturing information systems based on semantic entity," *International Journal of Smart Home*, vol. 10, no. 9, pp. 31–44, 2016.
- [20] L. Proskuryakova, D. Meissner, and P. Rudnik, "The use of technology platforms as a policy tool to address research challenges and technology transfer," *Journal of Technology Transfer*, vol. 42, no. 1, pp. 206–227, 2017.
- [21] X. Lin, J. Wang, and C. Lin, "Research on 3d reconstruction in binocular stereo vision based on feature point matching method," *Paper Asia*, vol. 2, no. 1, pp. 21–25, 2019.
- [22] X. Yang, M. Zhao, L. Shi, Z. Chen, and M. Zheng, "Research and implementation of fast algorithm for intra prediction mode selection oriented to hardware," *Journal of Computer-Aided Design and Computer Graphics*, vol. 31, no. 1, p. 158, 2019.