Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 521383, 7 pages
http://dx.doi.org/10.1155/2013/521383

Research Article

Robust People Tracking Using an Adaptive Sensor Fusion between a Laser Scanner and Video Camera

Yeong Nam Chae,[1] Yeong-Jae Choi,[1] Yong-Ho Seo,[2] and Hyun S. Yang[1]

[1] *Department of Computer Science, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea*
[2] *Department of Intelligent Robot Engineering, Mokwon University, 88 Doanbok-ro, Seo-gu, Daejeon, Republic of Korea*

Correspondence should be addressed to Yong-Ho Seo; yhseo@mokwon.ac.kr

Received 25 January 2013; Accepted 28 February 2013

Academic Editor: Sabah Mohammed

Robust detection and tracking in a smart environment have numerous valuable applications. In this paper, an adaptive sensor fusion method which automatically compensates for bias between a laser scanner and video camera is proposed for tracking multiple people. The proposed system comprises five components: blob extraction, object tracking, scan data clustering, a cluster selection, and updating the bias. Based on the position of object in an image, the proposed system determines the candidate scan region. Then, the laser scan data in the candidate region of an object is clustered into several clusters. A cluster which has maximum probability as an object is selected using a discriminant function. Finally, a horizontal bias between the laser scanner and video camera is updated based on the selected cluster information. To evaluate the performance of the proposed system, we show error analysis and two applications. The results confirm that the proposed system can be used for a real-time tracking system and interactive virtual environment.

1. Introduction

Robust detection and tracking in a smart environment have numerous valuable applications, such as recognizing human behavior for intelligent surveillance, monitoring, and analyzing. To monitor human activities such as location, identity, and behavior, robust detection and tracking are necessities in various environments.

In the people motion tracking area, the one side view, which comes from a single video image, is used to verify peoples' actions. In diverse and sophisticated environments, there are numerous problems in a single video image. In ubiquitous environments, to perceive the motion of people, unfavorable conditions exist in single-sensor systems, such as illumination variation and shadow. To overcome these problems, many methods rely on fusing a number of sensors, such as the infrared cameras, laser range finders, and image cameras of many directions [1, 2]. However, as various sensors are added, the object calibration is a very important issue in the object tracking area for the reliable detection and tracking of multiple objects [3].

To calibrate multi-sensors, an attempt using an extrinsic calibration between the camera and the laser scanner was proposed in [3–6]. This approach uses a technique in which both sensors image a planar checkerboard target at unknown orientations. Even if the calibration is reliable, it is inconvenient to adjust for tracking people simultaneously. Furthermore, the checkerboard always should be used to calibrate this system.

To overcome these problems, we present an adaptive sensor fusion method between the laser scanner and video camera. In this proposed approach, our system does not need a checkerboard. As the configuration of the system, which consists of a laser scanner and video camera, changes to intuitive positions, the traditional system requires a recalibration process. The proposed system, which has a background model, can compensate these sensor's horizontal variations automatically.

The paper is organized as follows: in the next section, we briefly introduce the proposed system. In Section 3, we present a video processing method for tracking multiple

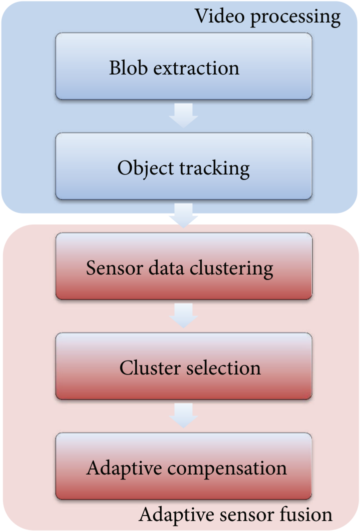


Figure 1: System overview.

objects. In Section 4, we present a detailed method to calibrate laser scanner and video camera. Then, we present the results with real environmental data in Section 5. At last, we conclude in Section 6.

2. System Overview

The proposed system consists of the video processing and the adaptive sensor fusion as shown in Figure 1. To measure the position of objects in an image, we adopt Mixture of Gaussians-(MoGs-) [7] based blob extraction and inference-graph-based object tracking approach. Then laser, scan data in the candidate region from the result of the object tracking is clustered into several clusters. A cluster which has maximum probability as an object is selected using a discriminant function. Finally, a horizontal bias between the laser scanner and video camera is compensated based on selected cluster information.

3. Video Processing

In order to merge the laser scanner and video camera, the proposed system performs video processing in a prior part of the system. In order to measure the position of objects in an image, we developed an enhanced view-based multiple-object tracking system based on previous research [8, 9]. In this section, we briefly introduce this multiple-object tracking system.

3.1. Blob Extraction. For segmentation, MoGs is widely used due to the capability for adaptation to the various environmental changes like illuminations. But, the main problem of MoGs is that moving objects are learned as background when they stop. And thus it fails to segment objects stopped as foreground. In the proposed system, in order to segment an input image into a foreground and background, a modified MoGs [8] is used to keep moving objects segmented as foreground, even when they stop. After conventional MoGs is performed, the following four steps are further executed to manage objects stopped. First, the Gaussians are sorted in descending order of their weight. Second, the pixels moving from the first Gaussian component to the second Gaussian component are identified as pixels belonging to the objects stopped and they are put into an augmented mask. Third, the pixels in the augmented mask are added to the segmentation mask from the conventional MoGs. Fourth, the pixels in the augmented mask are removed from the augmented mask when the stopped objects start to move. This modified MoGs guarantees that the objects still identified as foreground, even when moving objects stopped. In order to remove shadows and highlights, we can adjust the intensity of a shadow pixel compared with a background model as shown in Figure 2(d). After removing shadows and highlights, when capturing foreground pixels at frame t, they are clustered into a set of $B^t = \{b_i^t \mid i \text{ is an integer and } 0 \leq i\}$, where a blob b_i^t is an ith set of connected foreground pixels.

3.2. Object Tracking. A blob represents an object in the ideal case. In the real environment, however, one object can have several blobs (fragmentation), and one blob can have multiple-objects (grouping). To deal with these problems, Choi et al. [9] adopt an online multiple object tracking framework. Figure 3 shows the overall procedure of this framework. First, detecting blob association events between B^t and B^{t-1} can update the blob inference graph, labeling each vertex as fragment, object, and group. Finally, localization of objects can be captured by using the blob graph.

4. Adaptive Sensor Fusion

In this paper, we propose an adaptive method which compensates horizontal bias between the laser scanner and video camera. In order to merge the laser scanner and video camera adaptively, there are three steps in the proposed method. First, laser scan data in the candidate region of the object is clustered into several clusters. Second, a cluster which has maximum probability as an object is selected using a discriminant function. Finally, a horizontal bias between the laser scanner and video camera is updated based on selected cluster information.

4.1. Sensor Data Clustering. To match an object in an image to laser scan data, we determine the candidate region in the laser scan data. The candidate region in the laser scan data can be determined based on the object position in an image. This candidate region (θ_l, θ_r) is shown in the following equations:

$$\begin{aligned}\theta_l &= (1-\varphi)\left(90 + \text{fov}\left(-\frac{1}{2} + \frac{o_l}{\omega}\right) + \text{bias}\right),\\ \theta_r &= (1+\varphi)\left(90 + \text{fov}\left(-\frac{1}{2} + \frac{o_r}{\omega}\right) + \text{bias}\right).\end{aligned} \tag{1}$$

In (1), θ_l and θ_r are the left and right angle boundaries, respectively, o_l denotes the left pixel position of an object, o_r denotes the right pixel position of an object, ω refers to image

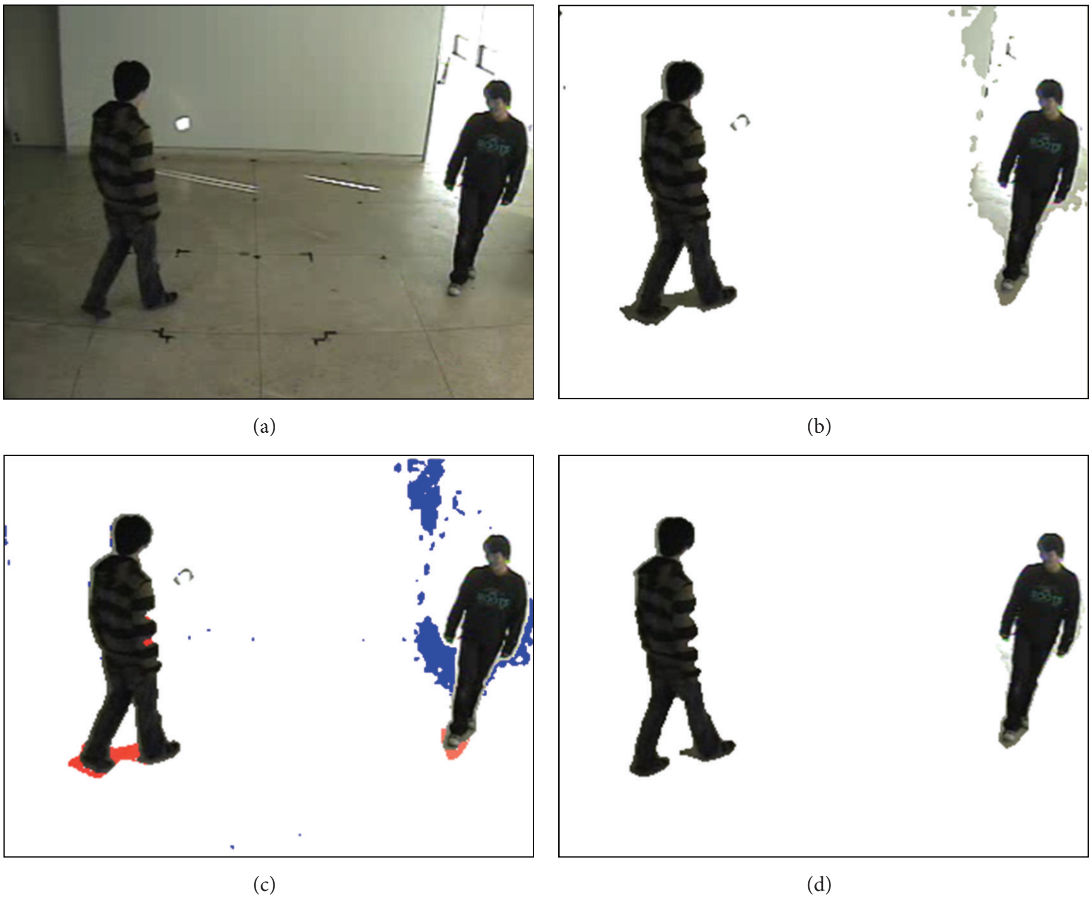


Figure 2: (a) An input image. (b) The result of foreground extraction. (c) Detected shadow (marked in red) and highlight (marked in blue). (d) The result of foreground extraction: an example of a figure without shadow or highlight.

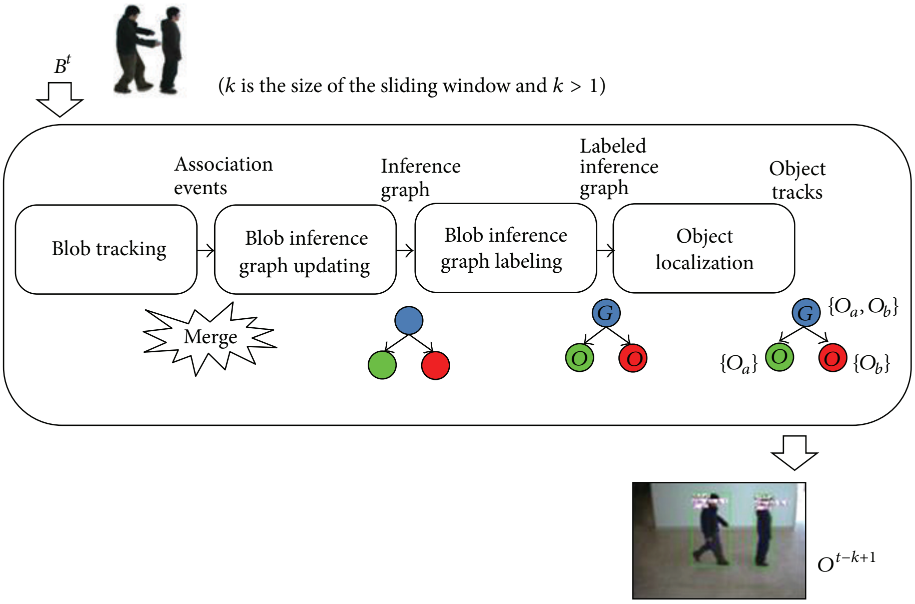


Figure 3: The online multiple-object tracking framework.

width, bias means horizontal bias between the laser scanner and video camera, fov means field of view angle, which is determined by the video camera, and φ denotes the scale factor which expands the candidate region.

The archived laser scan data in the candidate region is lumped into several clusters by the nearest neighborhood clustering algorithm. We use the following equation as a threshold equation:

$$\sqrt{d_{\mathrm{MV}}(\theta_1)^2 + d_{\mathrm{MV}}(\theta_2)^2 + 2d_{\mathrm{MV}}(\theta_1)\, d_{\mathrm{MV}}(\theta_2)\cos(|\theta_1 - \theta_2|)} > \delta. \tag{2}$$

In (2), $d_{\mathrm{MV}}(\theta)$ refers to the distance from the measured value of the laser scanner at angle θ, both θ_1 and θ_2 are discrete angle values $(0, 1, 2, \ldots 180)$, and δ denotes the neighborhood distance threshold value, which is set as 2.5 feet (an average person's stride length [10]). Using this threshold function, we classified the datum in 2.5 feet from each other as same cluster.

4.2. The Cluster Selection Method. In order to take an appropriate cluster which has the greatest probability to be an object, the proposed system adapts the following discriminant function:

$$m = \underset{0 \le i \le k}{\arg\max} \sum_{\theta \in \mathbf{C}_i} \Delta(\theta). \tag{3}$$

In (3), k denotes the number of clusters, and C_i refers to the ith cluster. A discriminant function Δ is adapted to compare each cluster:

$$\Delta(\theta) = \begin{cases} \dfrac{\theta_\omega}{2}\left|\theta - \theta_c\right| d_{\mathrm{MAX\,BG}}, & d_{\mathrm{BG\,MV}}(\theta) > \delta, \\[2ex] \dfrac{\theta_\omega}{2}\left|\theta - \theta_c\right| d_{\mathrm{BG\,MV}}(\theta), & \text{otherwise}. \end{cases} \tag{4}$$

In (4), $d_{\mathrm{MAX\,BG}}$ denotes the maximum distance from the background model of the laser scan data, $d_{\mathrm{BG\,MV}}(\theta)$ denotes the difference between the background model and measured value at θ, θ_ω refers to angular width, and θ_c refers to the center angle of the candidate region (θ_l, θ_r). These are shown in the following equations:

$$d_{\mathrm{MAX\,BG}} = \max_{0 \le \theta \le 180} d_{\mathrm{BG}}(\theta), \tag{5}$$

$$d_{\mathrm{BG\,MV}}(\theta) = \left|d_{\mathrm{BG}}(\theta) - d_{\mathrm{MV}}(\theta)\right|, \tag{6}$$

$$\theta_\omega = \left|\theta_l - \theta_r\right|, \tag{7}$$

$$\theta_c = \frac{\theta_l + \theta_r}{2}. \tag{8}$$

In (5) and (6), $d_{\mathrm{BG}}(\theta)$ denotes the distance from the background model of the laser scan data at θ. The background model of the laser scan data is similar to the traditional background model of the gray scale video stream. Consider

$$D_{\mathrm{BG}}(\theta) = \begin{cases} d_{\mathrm{BG}}(\theta) + \alpha\left(d_{\mathrm{MV}}(\theta) - d_{\mathrm{BG}}(\theta)\right), & d_{\mathrm{BG\,MV}}(\theta) \ge \delta, \\ d_{\mathrm{BG}}(\theta) + \beta\left(d_{\mathrm{MV}}(\theta) - d_{\mathrm{BG}}(\theta)\right), & \text{otherwise}. \end{cases} \tag{9}$$

In (9), α refers to the learning rate in the case of foreground and β refers to the learning rate in the case of the background. Equation (9) is adapted to the whole range of laser scan data.

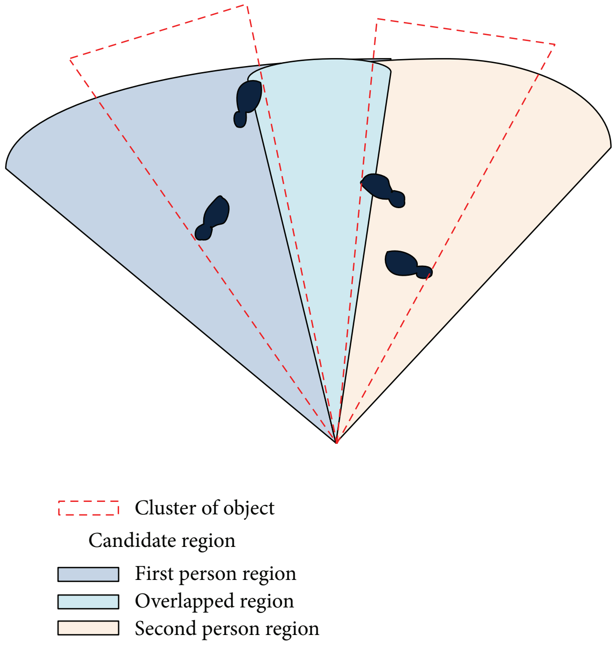


FIGURE 4: Example case of overlapping candidate regions.

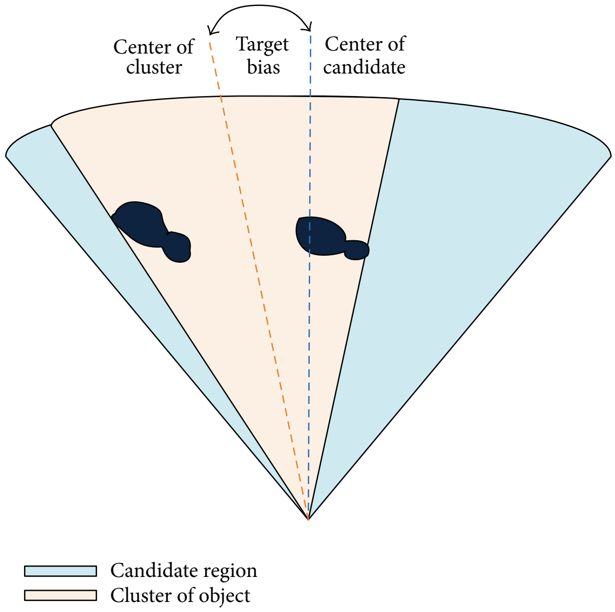
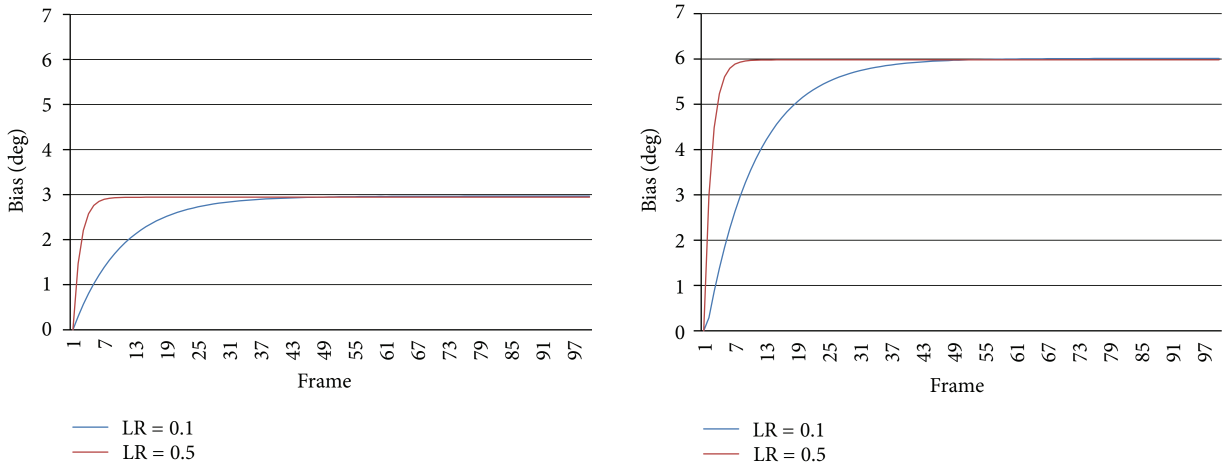


FIGURE 5: The horizontal bias angle.



(a) Degree 3

(b) Degree 6

Figure 6: The example of bias.

According to the discriminant function, the cluster volume, difference from the background model, and difference from the center of candidate region are considered as selection criteria. Using these selection criteria, the proposed system can select the appropriate cluster in case the candidate regions overlap. An example case of overlapping is shown in Figure 4.

4.3. Updating Rule. After determining an appropriate cluster, which is selected by a discriminant function, the proposed system updates bias adaptively. Figure 5 shows the target bias which is calculated from the candidate region and selected cluster. In order to take an appropriate cluster which has the greatest probability to be an object, the proposed system adapts the following discriminant function.

To diminish the target bias, the proposed system updates the bias using the following updating rule:

$$\text{bias} = \text{bias} + \rho \left(\text{MidAng}(m) - \theta_c \right). \tag{10}$$

In (10), bias refers to the horizontal bias angle between the field of view angle and the matched cluster, ρ denotes the learning rate, and MidAng(m) means the median angle of matched cluster m. The MidAng is calculated by the following equation:

$$\text{MidAng}(i) = \frac{\left| \min_{\theta \in \mathbf{C}_i} \theta + \max_{\theta \in \mathbf{C}_i} \theta \right|}{2}. \tag{11}$$

Using (10) as an updating rule, the proposed approach can compensate for horizontal bias adaptively.

5. Experimental Results

In order to evaluate proposed system, we arrange the video camera and the laser scanner vertically. Samsung SDC-415A model is adopted as video camera. It supports 768 × 494 resolution and covers 140 degree fields of view. SICK LMS100 model is adopted as laser scanner. It supports a 50 Hz scan rate over 270 degree range and 0.25 degree angular resolution. Its sensing range is 18 meters with an error of about 20 mm. The experimental results consist of two parts: the error analysis of the proposed approach and the application of the proposed system.

Table 1: The relative bias after convergence.

Learning rate	Angle			
	6	3	−3	−6
0.1	6.013	2.951	−3.139	−6.178
0.5	5.981	2.986	−3.080	−6.186

5.1. Error Analysis. To evaluate the proposed approach, we performed an error analysis of the bias compensation. To measure the error of the compensation, we installed a marker at the same position which can be detected by both the vision and laser scanner. By rotating the video camera three degrees horizontally, we measured the relative bias from the origin angle after convergence. The overall results are shown in Table 1. In this experiment, measurements were taken five times and the results were averaged out.

As shown in Table 1, the average error of the proposed approach is about 0.085 degree. The saturation of bias as time is shown in Figure 6. In Figure 6, LR refers to learning rate. In the case of 0.1, the saturation time is about 40 frames. However, in the case of 0.5, the saturation time is lower than 10 frames. From these results, the proposed approach can be considered reasonable for real-time systems.

5.2. Application. In this subsection, we show two applications of the proposed system. The first application is people tracking. The second application is a virtual pet system using augmented reality.

Figure 7 shows an example of multiple people tracking. The upper two images illustrate that the candidate regions do

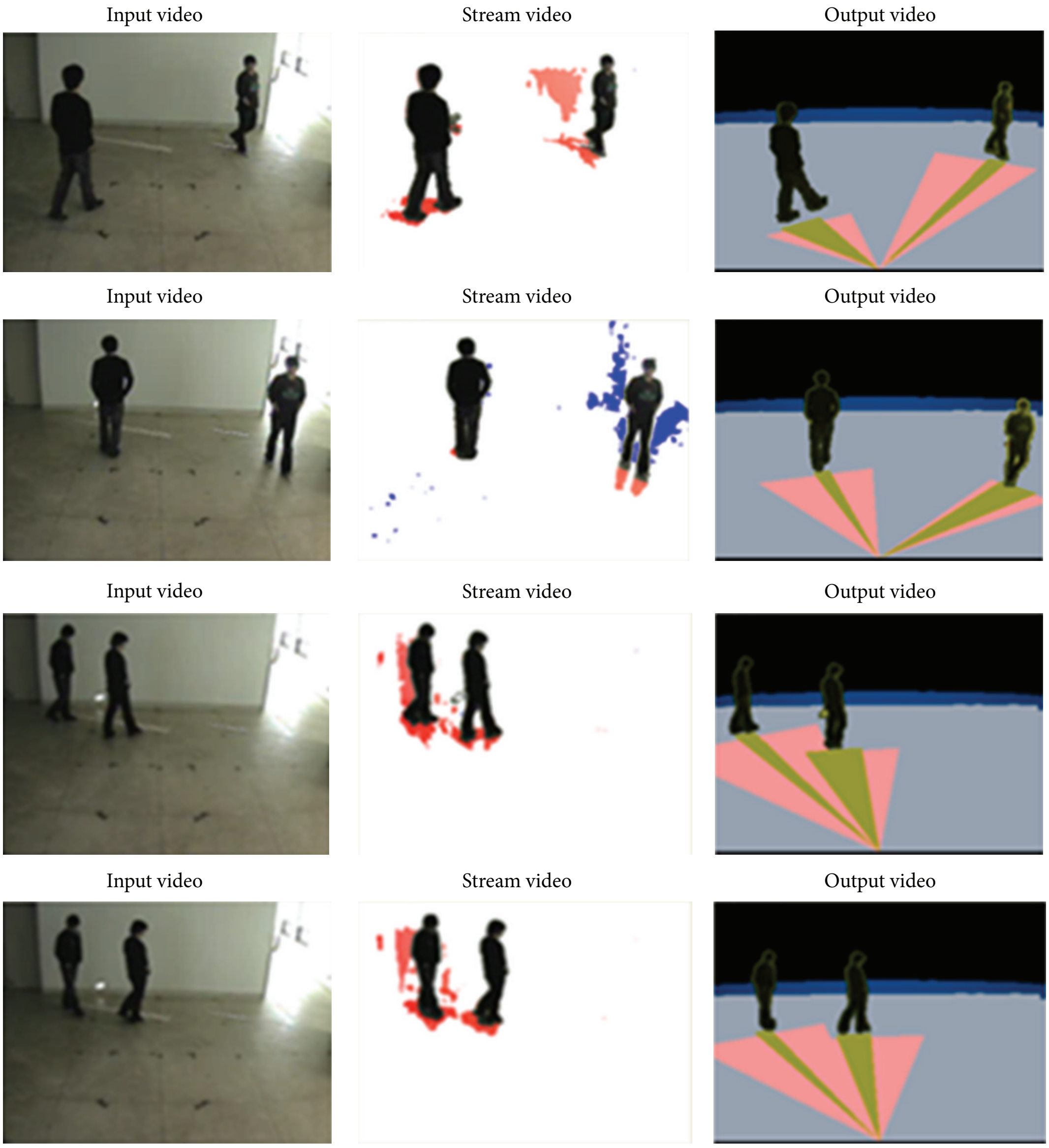


Figure 7: The people tracking by adjusting adaptive calibration.

not overlap, and the lower two images illustrate overlapping case. In both conditions, the proposed system can track appropriate people by the discriminant function.

In the second application, to apply the proposed system, we employed it in a virtual pet system based on augmented reality. As shown in Figure 8, only a person and a dog's house exist in a real environment. With the position of the calibrated object result and the human action recognition result, user can make a relationship with a virtual pet, called Cho-Rong-I.

The ground of the real environment is calibrated in the initial stage so that the bottom-center coordinate of each object can be converted into the coordinate in the ground.

We made several complex scenarios in order to probe the performance of the proposed system as follows: Cho-Rong-I follows the owner in Figures 8(a) and 8(b) and pretends to die when the owner pretends to shoot in Figure 8(c). Figures 8(d) to 8(f) show that Cho-Rong-I passes under the owner's legs. As the proposed system obtained the precise position, the augmented dog, Cho-Rong-I, also reacts with the person at the virtual region matching the real coordinates.

6. Conclusion

In this paper, we proposed adaptive sensor fusion methods that compensate for horizontal bias between a laser scanner and video camera for tracking people in real-time system. The usual method using the checkerboard is inconvenient to track people simultaneously because of the manual calibration in a previous research. The proposed system in this paper overcomes the problem using an automatic adaptive sensor fusion method in real-time people tracking. In this application field, the accuracy of compensation is a significant factor for a

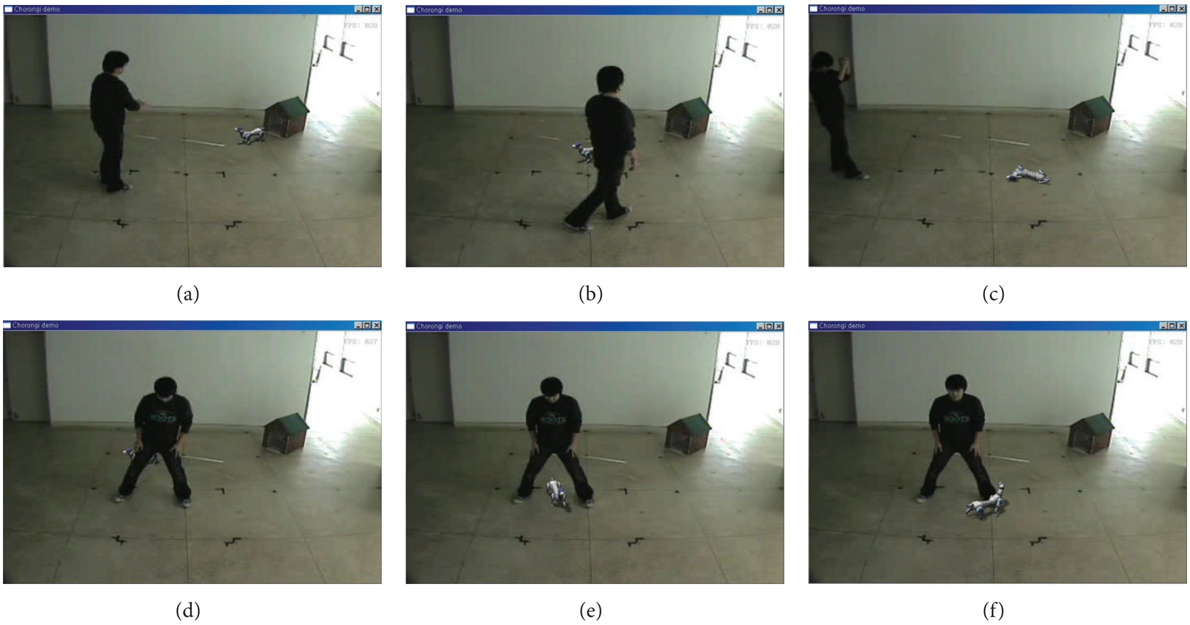

(a) (b) (c) (d) (e) (f)

Figure 8: Demonstration of the virtual pet system.

real-time system. In order to match images between the video camera and laser scanner, we propose the algorithm to merge the laser scanner and video camera simultaneously by capturing the position of the candidate region and the cluster of the laser scan datum. To evaluate the performance of the proposed system, we employed it for tracking two people and then applied it in an augmented reality where one person can interact with a virtual pet. These results show that the proposed system can be successfully employed to obtain the peoples' position by automatic sensor fusion. To enhance the proposed system, variations of the other axis should also be considered. We are currently improving our system to include the uncertainty variations of the sensor position.

Acknowledgments

This work was supported by the Basic Science Research Program through a Grant from the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0005793 and 2011–0013776). This work was also supported by the IT R&D Program of MKE/KEIT "Development of learner-participatory and interactive 3D virtual learning contents technology" (10039165) and the NAP (National Agenda Project) of Korea Research Council of Fundamental Science and Technology.

References

[1] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Multi-modal tracking of people using laser scanners and video camera," *Image and Vision Computing*, vol. 26, no. 2, pp. 240–252, 2008.

[2] Y. Bok, Y. Hwang, and I. S. Kweon, "Accurate motion estimation and high-precision 3D reconstruction by sensor fusion," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '07)*, pp. 4721–4726, April 2007.

[3] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS '04)*, pp. 2301–2306, October 2004.

[4] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[5] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, pp. 666–673, September 1999.

[6] C. Mei and P. Rives, "Calibration between a central catadioptric camera and a laser range finder for robotic applications," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '06)*, pp. 532–537, May 2006.

[7] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, pp. 246–252, June 1999.

[8] Y. Tian, A. Senior, and M. Lu, "Robust and efficient foreground analysis in complex surveillance videos," *Machine Vision and Applications*, vol. 23, no. 5, pp. 967–983, 2012.

[9] J. Choi, Y. Cho, K. Cho, S. Bae, and H. Yang, "A view-based multiple objects tracking and human action recognition for interactive virtual environments," *International Journal of Virtual Reality*, vol. 5, no. 3, pp. 1–6, 2006.

[10] P. Ryan, "Count the steps to motivate walking," *Functional U*, vol. 3, no. 3, pp. 12–16, 2005.