# Human Tracking with Multiple Cameras Based on Face Detection and Mean Shift

Atsushi Yamashita, Yu Ito, Toru Kaneko and Hajime Asama

***Abstract*— Human tracking is an important function to an automatic surveillance system using a vision sensor. Human face is one of the most significant features to detect person(s) in an image. However, face is not always observed from a single camera. Therefore, it is difficult to identify a person exactly in an image due to the variety of poses. This paper describes a method for automatic human tracking based on the face detection using Haar-like features and the mean shift tracking method. Additionally, the method increases its trackability by using multi-viewpoint images. The validity of the proposed method is shown through experiment.**

## I. INTRODUCTION

In this paper, we propose a multi-viewpoint human tracking system based on the face detection based on Haar-like features and the mean shift tracking method. The main contribution of the paper is to introduce multiple viewpoints to the video content indexing method [1] for improving robustness and accuracy of human tracking.

In recent years, surveillance by using cameras is essential to the civil life for safety and security. Research and development of automatic surveillance systems are now one of the most active topics because human resources are limited. Actually, surveillance camera systems based on computer vision techniques are widely used owing to the performance improvement and the cost reduction in computers and image input devices.

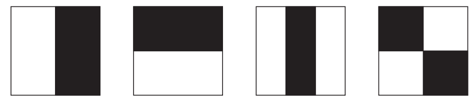
One of the main objectives of automatic surveillance camera systems is the analysis of human behaviors. For example, locations of each person can be found by recording the trajectories of individuals, warnings can be given to persons who are approaching to dangerous areas by sensing proximity, and so on. In these cases, human tracking is essential function to surveillance systems.

Before human tracking is executed, a region extraction that distinguishes human from other objects must be done at first. Once human is detected in an image, the detected human is tracked in an image sequence.

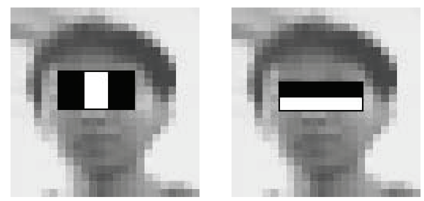
Human face is one of the most significant features to detect person(s) in the image. Therefore, there are a lot of studies about human face detection [2]–[6]. Previous studies can be divided into two approaches; heuristics based methods and machine learning based methods.

Atsushi Yamashita and Hajime Asama are with Department of Precision Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan {yamashita, asama}@robot.t.u-tokyo.ac.jp

Yu Ito and Toru Kaneko are with Department of Mechanical Engineering, Shizuoka University, 3–5–1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka 432–8561, Japan



(a) Examples of Haar-like features.



(b) Face detection.

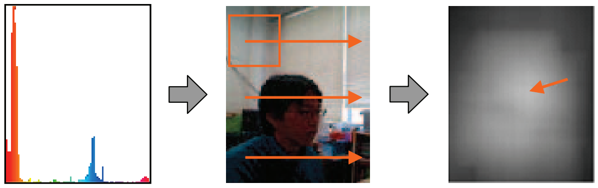
Fig. 1. Face detection by using Haar-like features.

Heuristic based approaches utilize characteristics of the face such as edge, skin color, hair color, alignment of face parts, and so on. For example, Govindaraju *et al.* use a deformable face template [7], and Sun *et al.* use color and local symmetry information to detect human faces [8].

The latter approaches solve the face detection problem as the clustering problem into two classes (face and non-face) by using machine learning algorithms. Viola and Jones propose a rapid face detection method based on a boosted cascade of simple feature classifiers [9]. Lienhart and Maydt develop Viola and Jones method [10]. These methods use Haar-like features [11] and AdaBoost algorithm [12] to achieve fast and stable face detection. Haar-like feature (Fig. 1) is one of the powerful tools of the object detection, not only for face detection, but also car detection and so on [13], [14].

Usually, one classifier corresponds to one face orientation. Therefore, the classifier for erecting frontal faces can not detect tilted frontal faces and side faces. In other words, it is difficult to treat the face rotation in plane and the face rotation out of plane. A face orientation in acquired images from the surveillance camera is not constant and changes constantly in contrast to a frontal face in commemorative photos. In this case, successful rate of face detection declines.

Therefore, it is desirable for stable human tracking to utilize another approach in addition to the face detection. For example, mean shift [15] and CONDENSATION [16] can track objects even when their orientations, shapes, and sizes changes. These tracking methods works well for the human tracking [1], [14], [17]–[19]. They are robust against shape changes of targets or partial occlusions. For example, Chateau *et al.* uses Haar-like features to detect faces and



(a) A color histogram is made as tracking model.

(b) Similarity distribution is calculated from (a) and color histogram in image.

(c) The position of a tracking object is searched by mean-shift.

Fig. 2. Mean shift tracker.

CONDENSATION algorithm [16] to track them [14].

Bradski [17] and Comaniciu *et al.* [18] proposed methods to track objects using mean shift (Fig. 2). These methods make color histogram of rectangular region in image. Then, similarity distribution is calculated from color histogram of rectangular region and tracking model. They search for tracking objects from similarity distribution. If color histograms are similar to tracking model, these methods can track objects. Accordingly, these methods have advantages that it is robust against the partial occlusions and shape changes of tracking objects.

For a video content indexing, Jaffre and Joly propose a method that detect faces in image by using Haar-like features [10] and then track their costumes by using mean shift algorithm [1]. This method is for TV talk-shows, and therefore targets usually face the camera. However, in the case of the surveillance system, back shots are not rare. These methods sometimes fail if there is only a back shot image.

On the other hand, there are a lot of studies about multi-viewpoint in image recognition technology. The purpose of multi-viewpoint is improvement of detection accuracy. This is based on the idea that objects are easy to be detected from various angles by multi-viewpoint. In real world applications, a multi-viewpoint approach is realistic and reasonable one [20]–[22].

## II. Our Approach

The face detection method using Haar-like features and classifiers made by statistical learning can work well under the complicated background. Therefore, visual surveillance systems using this method do not have a large problem in detection accuracy.

However, if surveillance systems track humans by using only this method, it is insufficient. In case surveillance systems detect faces, an appearance of face changes by angle and direction of tracking object. This method can detect faces only from limited viewpoint(s). For example, classifiers of front faces can not detect side faces. In case of surveillance, surveillance systems must detect faces in arbitrary postures. Therefore, human tracking systems using only this method can not work well in several situations. To overcome this problem, our method uses a mean shift tracker.

The mean shift tracker is the technique to track objects. This method has an advantage that it is robust against the change of tracking objects. However, the object model that is prepared beforehand is necessary for tracking. Therefore, our method uses color histogram of a face area that is detected by Haar-like features as object model.

Therefore, we propose a human tracking method based on face detection using Haar-like features and a mean shift tracker. However, there is a problem that tracking never starts until a face is detected initially by using Haar-like features. It takes time to detect a front face by this method using a single viewpoint.

Therefore, we introduce multiple viewpoints to solve this problem. The main contribution of the paper is to introduce multiple viewpoints to the human detection method [1] for improving robustness and accuracy of human tracking.

## III. Outline of Our Method

The flow of the human tracking process of a single viewpoint is described as follows (Fig. 3).

Our method tracks human by face detector using Haar-like features and a mean shift tracker. A front face area of a human is detected by the face detector using Haar-like features in an input image. If the detected human has not yet been registered, the detected human is registered newly as a tracking object. If the detected human has been already registered, the position and the size of the face area of the detected human are updated. When the position of the face area of the detected human exists near the position of tracked human in past image, the detected human has been already registered. If the front face area of the tracking human is not detected, it is tracked by the mean shift tracker. After the human tracking processing of a single viewpoint finished, the human tracking processing of multi viewpoint starts.

The flow of the human tracking process of a multi-viewpoint is described as follows (Fig. 4).

If a front face is detected by the human tracking process of a single viewpoint, the information of a position and a size of a front face area are sent to other viewpoints. When the information of a position and a size of a front face area are received from other viewpoints, the area that corresponds to a front face is searched by the information. In this paper, the area that corresponds to a front face is the back part of the head of the same human. Then, the back part of the head that is detected by the information is registered newly. The detected area of the back part of the head is tracked from next image in each viewpoint. Also, after the back part of the head is detected, the false-correspondence is diminished.

## IV. Human Tracking from a Single Viewpoint

### A. Face Detection

A front face area is detected by the face detection method using Haar-like features [10] in input image.

Our method detects human faces from almost-totally frontal direction, left frontal direction, and right frontal direction by using three face detection classifiers (almost-totally frontal face detector, left frontal face detector, and

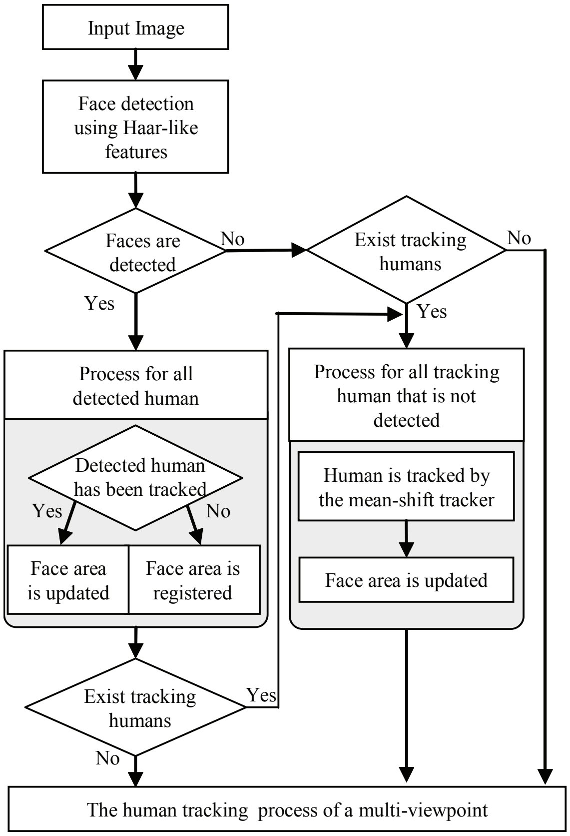



Fig. 3. A human tracking process of a single-viewpoint.

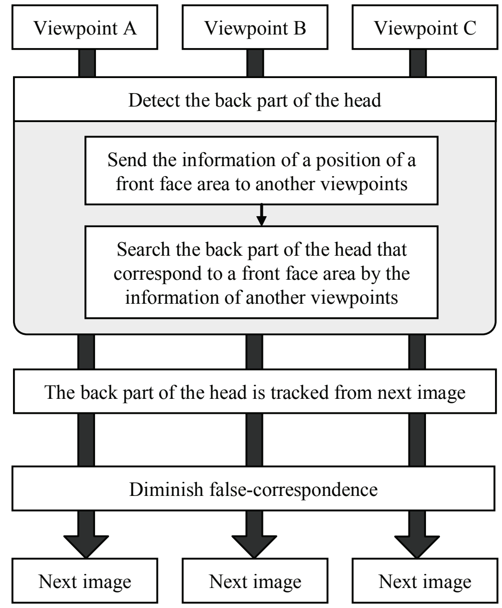
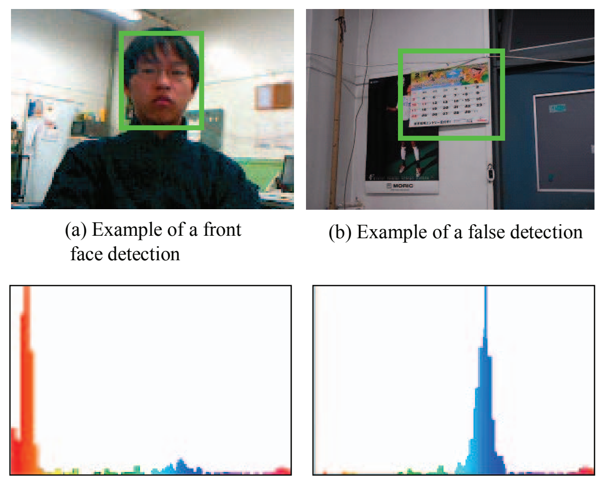



Fig. 4. A human tracking process of a multi-viewpoint.



(a) Example of a front face detection

(b) Example of a false detection

(c) Hue histogram of a front face area

(d) Hue histogram of a false detection area

Fig. 5. Face detection.

right frontal face detector) to improve the detection rate and robustness[1].

The result of front face detection is shown in Fig. 5(a). The rectangle area is registered newly as a tracking object or updated. After a front face is detected, hue histogram of face area is calculated. When detected human is tracked by the mean shift tracker, hue histogram is used as color information of tracking object. Also, because face detection method using Haar-like features uses gray-scale image, it detects object that is similar to hue pattern of front face like Fig. 5(b). Many of hue histogram of false detections like Fig. 5(d) may not be similar to that of a front face area like Fig. 5(c). Therefore, when this method detects a candidate for front face, it examines principal ingredient of hue histogram of a candidate for front face. In case of hue histogram of front face, principal ingredient is skin color. Then, if principal ingredient of hue histogram of a candidate for front face is not skin color, a candidate for front face is removed. Therefore, detection accuracy is improved.

### *B. Mean Shift Tracker*

If a face is not detected by face detector using Haar-like features, the position of the detected human is not updated. Then, the human can not be tracked. In case the human can not be tracked by face detector using Haar-like features, the human is tracked by the mean shift tracker. The mean shift tracker searches for the area whose color information is similar to it of tracking objects. Similarity distribution is calculated from two hue histograms by the Bhattacharyya coefficient [23], [24]. One histogram is made from the front face area that is detected by face detection method using Haar-like features. Other histogram is made from the rectangle area in present image. The Bhattacharyya coefficient $\rho$ is defined as:

$$\rho = \sum_{u=1}^{m} \sqrt{p_u q_u}, \tag{1}$$

where $p_u$ is a normalized hue histogram of the rectangle area in present image, $q_u$ is a normalized hue histogram of

[1]Note that henceforce "front" means almost-totally front, left front, and right front.

(a) Haar-like.

(b) Haar-like + mean shift.

Fig. 6. Human detection and tracking results.

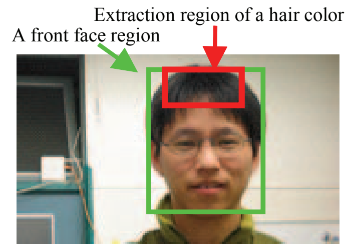
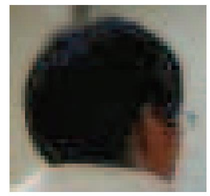
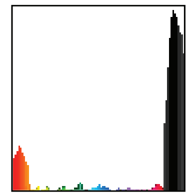



Fig. 7. Hair color estimation.



(a) Posture of expected back part of the head



(b) Hue histogram of the back part of the head

Fig. 8. Posture of expected back part of the head and its hue histogram.

object model, $u$ is the number of ingredient of hue histogram and $m$ is the total number of ingredient of hue histogram, respectively.

After similarity distribution is calculated by the above-mentioned procedure, our method searches for the area that is the most similar to color information of tracking objects by the mean shift tracker. Then, the position of face area is updated from this result.

Figure 6 shows face detection and tracking results. Figure 6(a) shows the result only with the face detection using Haar-like feature, and Fig. 6(b) shows the result both with the face detection and the mean shift tracker, respectively. A rectangle in image is a detected face area. Human tracking fails in middle image of (a), while it never fails in (b). From these results, it is verified that the combination of two methods can detect and track human robustly.

## V. Human Tracking from Multiple Viewpoints

### A. Extraction of Hair and Skin Colors

In the case of multiple viewpoints, a position of the same human in another camera is limited on an epilolar line by epipoplar constraints. However, the back part of a head of the same human can not always be detected by using only position information of a front face area.

Therefore, the hair color is estimated when the front face is detected from one viewpoint. Under the assumption that hair dominates the upper area of the detected face region, our method estimates hair color (Fig. 7).

Colors in this region are sorted in descending order of frequency, and mode value of the color is detected as a principle hair color. The distribution of hair color is decided by voting color values of each pixel in this region in the HSV color space.

After that, the back part of the head is detected and then it is tracked by using its hue histogram (Fig. 8).

In the same way, skin color is estimated by using the central area of the detected face region in this step.

### B. Correspondence Matching from Multiple Viewpoints

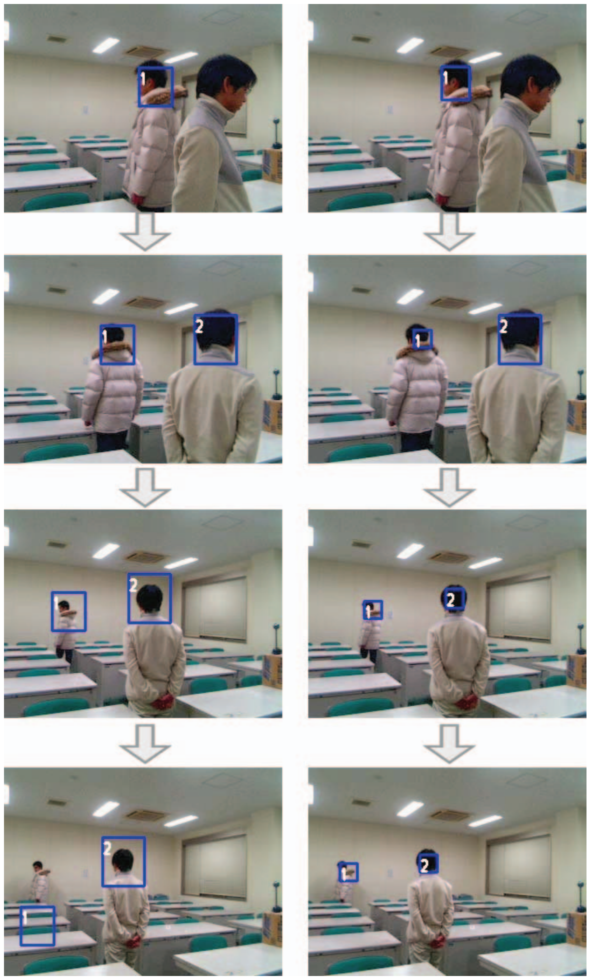
In the case of (more than) three-camera system, we can improve accuracy of tracking to reduce false-correspondence. Even when the back part of a head of the same human is detected by hair color information and epipolar constraints, there is possibility that a back part area except hair is detected. Thereupon, in case the same human is detected by each camera, the false-correspondence is diminished. To explain thereafter, we explain the case of three camera system and attach labels to three cameras. The camera that detects a front face is represented as camera (A). Then, the others are represented as camera (B) and camera (C).

At first, three-dimensional (3-D) coordinates are calculated from camera (A) and camera (B). Similarly, 3-D coordinates are calculated from camera (A) and camera (C). Then, two 3-D coordinates are compared. If two 3-D coordinates are the same, the back part of the head of the same human is detected correctly. If two 3-D coordinates are different, the uncolored area except hair is detected in camera (B) or/and camera (C).

If two 3-D coordinates are different, false-correspondence is found and corrected. Each 3-D coordinate is projected on the image that is acquired by another camera. Then, when the back part of the head of the same human exists in a projected area, projected 3-D coordinate is correct. When the back part of a head of the same human does not exists in a projected area, projected 3-D coordinate is incorrect. False-correspondence is corrected by the area where correct correspondence is projected.

### C. Adaptive Change of Tracker Size

The original camshift algorithm [17] can change tracker size (size of rectangle in Fig. 6) according to the target size in the image plane. In addition, our multi-view system can acquire 3-D positions of tracked persons. Therefore, 3-D positional information is utilized for changing the tracker



(a) Without tracker size change. (b) With traker size change.

Fig. 9. Adaptive change of tracker size (three viewpoints).

size. 3-D information makes the tracker size more accurate than only using image information (2-D information).

We can estimate the position of tracked human in time $t$ by using 3D coordinate values of the target in the past time $(t-1)$ and $(t-2)$.

Then, tracker size can be changed according to the distance between camera(s) and the target human(s).

$$s(t) = \frac{d(t-1)}{d(t)} s(t-1), \tag{2}$$

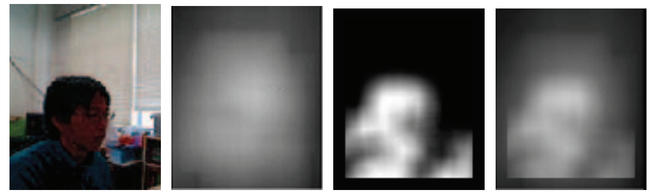
where $s(t)$ is the tracker size in time $t$, and $d(t)$ is the estimated distance between the camera and the target object in 3-D world coordinate, respectively[2].

Figure 9(a) shows the tracking result without a tracker size change. The tracker loses human-1. Figure 9(b) shows the successful result with our tracker size change method.

### D. Adaptive Change of Hair Color Distribution

Our tracking method using hue histogram is based on [17]. When this method tracks faces, the hue of the area whose saturation and brightness are low is ignored. Because the area whose saturation and brightness are low is easy to be influenced by lighting, they become unstable factors in object tracking.

[2]Initial value of tracker size can be determined from face detection results using Haar-like features.



(a) Image. (b) $\rho$. (c) $\lambda$. (d) $\gamma$.

Fig. 10. Share distribution ($\alpha = 0.7$).

In case tracking object is the back part of the head of detected human whose hair is uncolored, this method can not track it by the mean shift tracker. Our method uses the area whose color is detected in upper area of the front face image. In this way, our method can track the back part of the head of detected human.

However, the original tracking method [17] sometimes fails when the ratio of hair occupation area in the detected rectangular region changes while tracking. Therefore, we consider the hair occupation rate and redefine the color distribution defined in Eq. (1).

The hair occupation rate can be calculated by Eq. (3).

$$\lambda = \frac{p_f + p_h}{\sum_{u=1}^{m} p_u}, \tag{3}$$

where $p_u$ is a normalized hue histogram of the rectangle area, $p_f$ is a normalized hue histogram of skin color in the rectangle area, $p_h$ is that of hair color in the rectangle area, $u$ is the number of ingredient of hue histogram, $f$ is the number of ingredient of hue histogram of skin color, and $h$ is the number of ingredient of hue histogram of hair color, respectively.

After obtaining the color distribution $\rho$ (Eq. (1)) and the hair occupation rate $\lambda$ (Eq. (3)), we define the share distribution (Fig. 10) as composition of $\rho$ and $\lambda$.

$$\gamma = \alpha\rho + (1-\alpha)\lambda, \tag{4}$$

where $\alpha$ is a coefficient ($0 \leq \alpha \leq 1$).

We use the share distribution $\gamma$ in the mean-shift algorithm for robust tracking.

Figure 11(a) shows the tracking result by using the original color similarity distribution defined by Eq. (1), and Fig. 11(b) shows the result by using the new share distribution defined by Eq. (4), respectively. The tracker using the new share distribution continues to work well when the walking direction of human changes (Fig. 11(b)), while the tracker using the original color similarity distribution fails because the hair occupation rate changes (Fig. 11(a)).

From these results, it is verified that our method can track human robustly when the ratio of head in the detected rectangular region changes.

(a) Similarity distribution. (b) Share distribution.

Fig. 11. Adaptive change of hair color distribution (three viewpoints).

(a) Viewpoint (A). (b) Viewpoint (B).

Fig. 12. Experimental result (two viewpoints).

## VI. Experiment

Web cameras (Logicool Qcam Orbit MP QVR-13) were used in experiments and the image size was set as $320\times160$pixels.

The result of human tracking from two viewpoints (two cameras) is shown in Fig. 12. In this experiment, two cameras were set are face-to-face. Figure 12(a) is from viewpoint (A), and Fig. 12(b) is from viewpoint (B), respectively. Rectangles in image are detected face areas.

At first, the front face area is detected by face detection method using Haar-like features from viewpoint (A) and it is labeled human-1. The correspondence between two cameras is calculated and then human-1 is also detected from viewpoint (B), although the face of human-1 is not observed from viewpoint (B).

In the next frame, human-2 is detected and two humans can be tracked successfully regardless of their positions and poses.

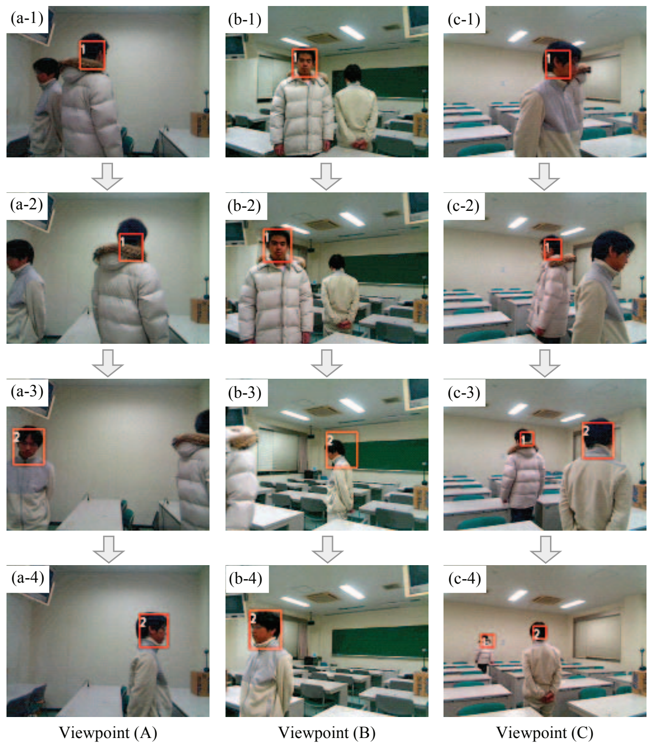
To verify the effectiveness of our method in the long term, experiments were done in the same room during any season.

A configuration of three cameras depends on environment. In this paper, three cameras are placed in triangle and an optical axis of each camera is turned to center of triangle. In this situation, the possibility that human can not be detected from three cameras decreases.

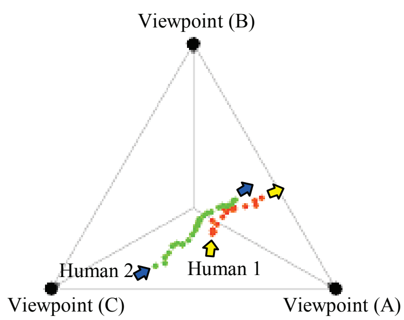
Figure 13(a) shows an example of the human tracking result from three viewpoints (three cameras) in winter, and Fig. 14(a) shows that in summer, respectively.

The front face area is detected by face detection method using Haar-like features in Fig. 13(b-1) and it is labeled human-1. The humans in Fig. 13(a-1) and Fig. 13(c-1) are detected by information of human-1 in Fig. 13(b-1). The detected human in Fig. 13(a-1) corresponds to human-1 in Fig. 13(b-1) correctly, but the detected human in Fig. 13(c-1) corresponds to human-1 in Fig. 13(b-1) incorrectly. However, false-correspondence in Fig. 13(b-1) is corrected in next image (Fig. 13(c-2)).

Next, the front face area is detected by face detection method using Haar-like features in Fig. 13(a-3) and it is labeled human-2. The humans in Fig. 13(b-3) and Fig. 13(c-3) are detected by information of human-2 in Fig. 13(a-3). Also, in case a front face is never detected by face detection method using Haar-like features in viewpoint (C), a face is detected by information of the other viewpoints. Then, detected human is tracked by the mean shift tracker.

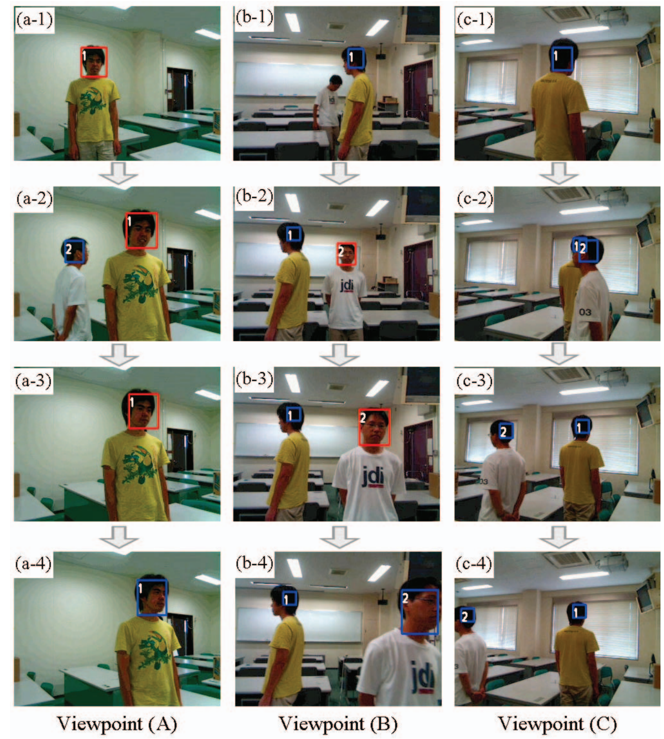



(a) Tracking result.

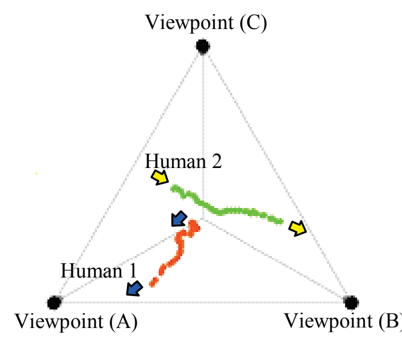



(b) Trajectories of humans.

Fig. 13. Experimental result in winter (three viewpoints).




(a) Tracking result.




(b) Trajectories of humans.

Fig. 14. Experimental result in summer (three viewpoints).

Figures 13(b) and 14(b) show trajectories of tracking humans. A trajectory of human-1 is represented as orange dots and a trajectory of human-2 is represented as green dots. Our method can track humans not only in images, but also in 3-D world coordinates.

Tables I and II show quantitative evaluation results.

Table I shows human detection accuracy of each camera. The results of data number 2 (camera 2), number 5 (camera 0), and number 8 (camera 2) are not so good. However, in these cases, other cameras can detect human successfully. Average detection rate is 97%. Therefore, from these results, it is verified that our method can detect human robustly.

Table II shows false-detection rate of each camera. The false detection occurs in data number 2 (camera 2), number 5 (camera 1), and number 6 (camera1). The tracker should disappear when humans are going outside of field of view. In these failure cases, the tracker still remained in images. However, our tracking system works well from the viewpoint of whole system in all cases.

As to computation time, the face detection using Haar-like features is 15fps, and the mean shift tracker is 20fps, respectively, when we use a single computer (CPU: Dual Core 1.6GHz, Memory: 2GB, OS: Windows XP). Computation time of a single viewpoint system is 10fps, and that of a three viewpoint system is 3fps, respectively. Generally speaking, our method can work well in real time.

These experimental results show the effectiveness of the proposed method.

## VII. Conclusion

In this paper, we propose a multi-viewpoint human tracking method based on face detection using Haar-like features

TABLE I

HUMAN DETECTION ACCURACY.

| Data number | Camera 0 | Camera 1 | Camera 2 |
|---|---|---|---|
| 1 | 99% | 100% | 96% |
| 2 | 100% | 100% | 70% |
| 3 | 100% | 100% | 100% |
| 4 | 100% | 100% | 100% |
| 5 | 86% | 100% | 100% |
| 6 | 100% | 100% | 100% |
| 7 | 100% | 100% | 100% |
| 8 | 100% | 78% | 100% |
| 9 | 100% | 100% | 100% |

TABLE II

FALSE-DETECTION RATE.

| Data number | Camera 0 | Camera 1 | Camera 2 |
|---|---|---|---|
| 1 | 3% | 0% | 0% |
| 2 | 0% | 0% | 21% |
| 3 | 0% | 0% | 0% |
| 4 | 0% | 2% | 0% |
| 5 | 6% | 58% | 0% |
| 6 | 0% | 52% | 0% |
| 7 | 0% | 0% | 0% |
| 8 | 1% | 12% | 1% |
| 9 | 0% | 0% | 0% |

and mean shift tracker. We confirm the effectiveness of the proposed method by experimental results.

The main contribution of the paper is to introduce multiple viewpoints to the human detection method [1] for improving robustness and accuracy of human tracking. We propose an automatic human head (hair) detection method, an adaptive tracking method for human tracking by improving the original mean-shift algorithm, and a reduction method of false-correspondence between multiple viewpoints.

As a future work, it is also important to reduce computation time by using multiple computers. This is because the computation time of our method increases as the number of humans under tracking increases. Human tracking with multi-sensor fusion approach is also effective to improve robustness [25]–[27].

## REFERENCES


[1] Gael Jaffre and Philippe Joly: "Costume: A New Feature for Automatic Video Content Indexing," *Proceedings of RIAO2004*, pp.314–325, 2004.

[2] Matthew A. Turk and Alex P. Pentland: "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol.3, No.1, pp.71–86, 1991.

[3] Matthew A. Turk and Alex P. Pentland: "Face Recognition Using Eigenfaces," *Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.586–591, 1991.

[4] Rama Chellappa, Charles L. Wilson and Saad Sirohey: "Human and Machine Recognition of Faces: A Survey," *Proceedings of the IEEE*, Vol.83, No.5, pp.705–741, 1995.

[5] Erik Hjelmas and Boon Kee Low: "Face Detection: A Survey," *Computer Vision and Image Uniderstanding*, Vol.83, pp.236–274, 2001.

[6] Ming-Hsuan Yang, David J. Kriegman and Narendra Ahuja: "Detecting Faces in Images: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.1, pp.34–58, 2002.

[7] Venu Govindaraju, Sargur N. Srihari, David B. Sher: "A Computational Model for Face Location," *Proceedings of the 1990 IEEE International Conference on Computer Vision*, pp.718–721, 1990.

[8] Qibin Sun, Weimin Huang, Jiankang Wu: "Face Detecion Based on Color and Local Symmetry Information," *Proceedings of the 3rd International Conference on Face and Gesture Recognition*, pp.130–135, 1998.

[9] Paul Viola and Michel J. Jones: "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.511–518, 2001.

[10] Rainer Lienhart and Jochen Maydt: "An Extended Set of Haar-like Features for Rapid Object Detection," *Proceedings of the 2002 IEEE International Conference on Image Processing*, Vol.1, pp.900–903, 2002.

[11] Constantine P. Papageorgiou, Michael Oren and Tomaso Poggio: "A General Framework for Object Detection," *Proceedings of the 6th IEEE International Conference on Computer Vision*, pp.555–562, 1998.

[12] Yoav Freund and Robert E. Schapire: "Experiments with a New Boosting Algorithm," *Proceedings of the 13th International Conference on Machine Learning*, pp.148–156, 1996.

[13] Kai She, George Bebis, Haisong Gu and Ronald Miller: "Vehicle Tracking Using On-Line Fusion of Color and Shape Features," *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems*, pp.731–736, 2004.

[14] Thierry Chateau, Vincent Gay-Belille, Frederic Chausse and Jean-Thierry Lapreste: "Real-Time Tracking with Classifiers," *Proceedings of ECCV2006 Workshop on Dynamical Vision*, 2006.

[15] Dorin Comaniciu and Peter Meer: "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No.5, pp.603–619, 2002.

[16] Michael Isard and Andrew Blake: "CONDENSATION - Conditional Density Propagation for Visual Tracking," *International Journal of Computer Vision*, Vol.29, No.1, pp.5–28, 1998.

[17] Gary R. Bradski: "Real Time Face and Object Tracking as a Component of a Perceptual User Interface," *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision*, pp.214–219, 1998.

[18] Dorin Comaniciu, Visvanathan Ramesh and Peter Meer: "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," *Proceedings of the 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.142–149, 2000.

[19] Dorin Comaniciu, Visvanathan Ramesh and Peter Meer: "Kernel-Based Object Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.25, No.5, pp.564–577, 2003.

[20] Robert T. Collins, Omead Amidi and Takeo Kanade: "An Active Camera System for Acquiring Multi-View Video," *Proceedings of the 2002 International Conference on Image Processing*, Vol.1, pp.527–520, 2002.

[21] Takekazu Kato, Yasuhiro Mukaigawa and Takeshi Shakunaga: "Cooperative Distributed Registration for Robust Face Recognition," *Systems and Computers in Japan*, Vol.33, No.14, pp.91–100, 2002.

[22] Kyungnam Kim and Larry S. Davis: "Multi-Camera Tracking and Segmentation of Occluded People on Ground Plane Using Search-Guided Particle Filtering," *Proceedings of the 9th Europian Conference on Computer Vision*, pp.98–109, 2006.

[23] Thomas Kailath: "The Divergence and Bhattacharyya Distance Measures in Signal Selection," *IEEE Transactions on Communication Technology*, Vol.COM-15, No.1, pp.52–60, 1967.

[24] Guorong Xuan, Peiqi Chai and Minhui Wu: "Bhattacharyya Distance Feature Selection," *Proceedings of the 13th International Conference on Pattern Recognition*, Vol.2, pp.195–199, 1996.

[25] Jinshi Cui, Hongbin Zha, Huijing Zhao and Ryosuke Shibasaki: "Tracking Multiple People Using Laser and Vision," *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.1301–1306, 2005.

[26] Ryo Kurazume, Hiroyuki Yamada, Kouji Murakami, Yumi Iwashita and Tsutomu Hasegawa: "Target Tracking Using SIR and MCMC Particle Filters by Multiple Cameras and Laser Range Finders," *Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.3838–3844, 2008.

[27] Taketoshi Mori, Takashi Matsumoto, Masamichi Shimosaka, Hiroshi Noguchi and Tomomasa Sato: "Multiple Persons Tracking with Data Fusion of Multiple Cameras and Floor Sensors Using Particle Filters," *Proceedings of ECCV2008 Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.