Asia International Symposium on Mechatronics 2008
27-31 August 2008 Hokkaido University, Japan

A Multi-Viewpoint Human Tracking Based on Face Detection Using Haar-like Features and a Mean-Shift Tracker

Yu Ito, Atsushi Yamashita and Toru Kaneko

Department of Mechanical Engineering, Shizuoka University, Shizuoka, Japan
(Tel : +81-53-478-1604; E-mail: {f0730011,tayamas,tmtkane}@ipc.shizuoka.ac.jp)

Abstract: Human tracking is an important function for an automatic surveillance system using a vision sensor. However, it is difficult to identify a human exactly in an image due to the variety of poses. This paper describes a method for automatic human tracking based on face detection using Haar-like features and mean-shift tracking. The method increases its trackability by using multi-viewpoint images. Experimental results showed the validity of the method.

Keywords: Human Tracking, Multi-Viewpoint, Haar-like Features, Mean-Shift Tracker.

1. INTRODUCTION

Visual surveillance systems using cameras increase in various environments by the rise of security needs. However, there is a limit in the manpower that can be used for surveillance. Therefore, researches are extensively done for automation of this work. In recent years, image recognition technology develops in a field of computer vision. Therefore, automated surveillance systems using visual information are becoming the effective means.

1.1 Background

The purpose of camera-based surveillance system is analysis of human actions. For example, the destination of human can be recorded by tracing a trajectory of human. Human tracking is necessary for analysis of human actions. Therefore, a human detection is necessary as preprocessing of human tracking.

In recent years, Viola and Jones[1], and Lienhart and Maydt[2] proposed the object detection method using Haar-like features (Fig. 1). This method can realize a rapid and stable object detection. It uses classifiers made by statistical learning[3]. In addition, because a face of human has a lot of information for surveillance, studies about face detection and recognition are proposed by other researchers [4][5].

Comaniciu *et al.*[6] and Bradski[7] 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 change of tracking objects.

Jaffre proposed study about video content indexing as

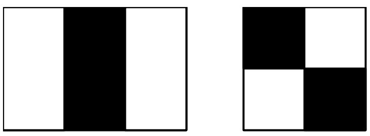
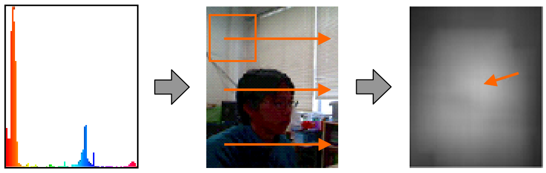


Fig. 1 Example of Haar-like features



(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

applied example of these methods[8]. After this method detected face by object detection method[2], it tracks costume by mean-shift tracker.

Also, 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.

1.2 Purpose

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 have no 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 viewpoints. 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.

2. OUTLINE

The flow of the human tracking process of a single viewpoint is described as follows (Fig. 3). Our method tracks human by face detection using Haar-like features and a mean-shift tracker. A front face area of a human is detected by the face detection 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 was detected by the human tracking process of a single viewpoint, the information of a position and a size of a front face area is sent to another viewpoints. When the information of a position and a size of a front face area is received from another 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 area of the back part of the head that is detected is tracked from next image in each viewpoint. Also, after the back part of the head is detected, the false-correspondence is diminished.

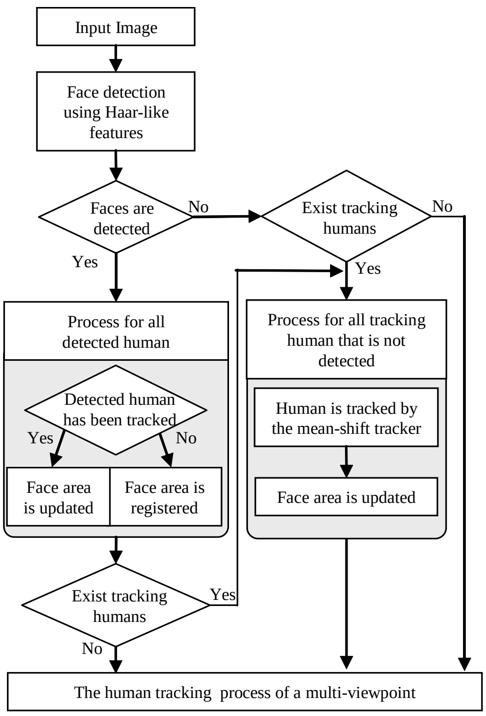


Fig. 3 A human tracking process of a single viewpoint

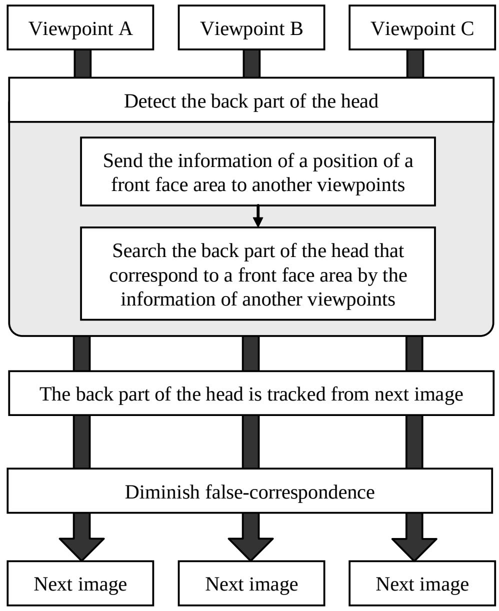


Fig. 4 A human tracking process of a multi viewpoint

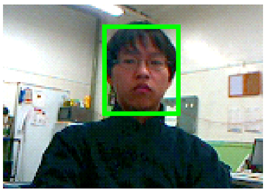
3. A HUMAN TRACKING PROCESS OF A SINGLE VIEWPOINT

3.1 Processing of face detection

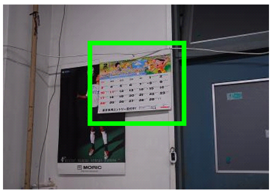
A front face area is detected by the face detection method using Haar-like features[2] in input image. The result of front face detection is shown in Fig. 5(a). This rectangle area is registered newly as a tracking object or updated. After a front face is detected, hue histogram of face area is calculated. 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.

3.2 Mean-shift tracker

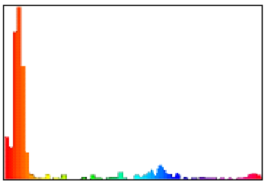
If a face is not detected by face detection using Haar-like features, the position of the detected human is not updated. Then, the human can not be tracked. In case



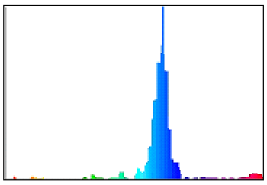
(a) Example of a front face detection



(b) Example of a false detection



(c) Hue histogram of a front face area



(d) Hue histogram of a false detection area

Fig. 5 Detected face area

the human can not be tracked by face detection 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[9]. 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 ρ is defined as:

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

where p_u is normalized hue histogram of the rectangle area in present image, q_u is normalized hue histogram of object model, u is the number of ingredient of hue histogram and m is the total number of ingredient of hue histogram, respectively.

Tracking method using hue histogram is based on [7]. 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. In case tracking object is the back part of the head of detected human whose hair is black, this method can not track it by the mean-shift tracker. Our method uses the area whose brightness is low as special color area. Therefore, our method can track the back part of the head of detected human whose hair is black.

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.

4. A HUMAN TRACKING PROCESS OF A MULTI-VIEWPOINT

4.1 Correspondence between multi-viewpoint

The human tracking process of a multi-viewpoint uses three cameras. A configuration of three cameras is different in environment. However, three cameras are placed in triangle and an optical axis of each camera is turned to center of triangle in this paper. Therefore, the possibility that human can not be detected from three cameras decreases.

When a front face area is detected by one camera, a position of the same human in the other cameras is presumed by epipoplar constraints. The back part of a head of the same human can not be detected by method using only position information of a front face area. However, when a front face area is detected by one camera, a posture of the same human in another camera can be expected from a location of three cameras. Then, a posture of the same human in another camera is the back part of the head from positions of three cameras. Our method has assumption that the color of a hair is black. The back part of a head of the same human is detected by this assumption and epipolar constraints.

4.2 Reduction of false-correspondence

In case the back part of a head of the same human is detected by this promise and epipolar constraints, there is possibility that a black 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 attach labels to three cameras. The camera detected a front face is represented as camera (A). Then, the others are represented as camera (B) and camera (C).

At first, three-dimensional coordinates are calculated from camera (A) and camera (B). Similarly, three-dimensional coordinates are calculated from camera (A) and camera (C). Then, two three-dimensional coordinates are compared. If two three-dimensional coordinates are the same, the back part of the head of the same human is detected correctly. If two three-dimensional coordinates are different, the black area except hair is detected in camera (B) or/and camera (C). If two three-dimensional coordinates are different, false-correspondence is found and corrected. Each three-dimensional 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 three-dimensional coordinate is correct. When the back part of a head of the same human does not exists in a projected area, projected three-dimensional coordinate is incorrect. False-correspondence is corrected by the area where correct correspondence is projected.

5. EXPERIMENT

The experimental environment is shown in Fig. 6. The result of human tracking using our method is shown in Fig. 7. A rectangle in image is a detected face area. Also, trajectories of tracking humans are shown in Fig. 8.

The front face area was detected by face detection method using Haar-like features in Fig. 7(b-1) and it was labeled human-1. The humans in Fig. 7(a-1) and Fig. 7(c-1) were detected by information of human-1 in Fig.

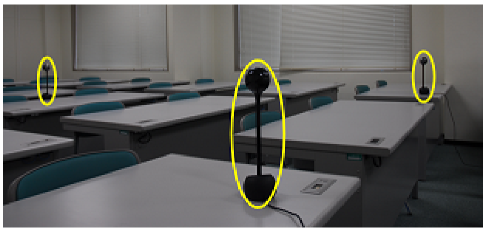


Fig. 6 The experimental environment

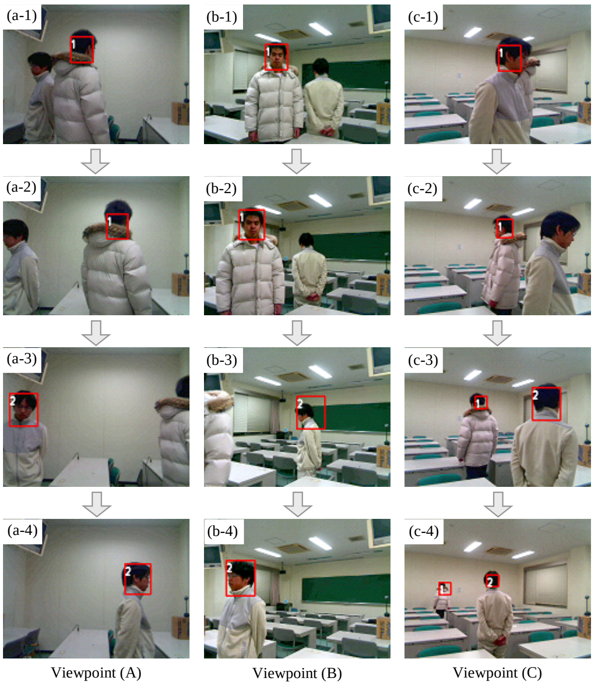


Fig. 7 Experimental result

7(b-1). The detected human in Fig. 7(a-1) corresponded to human-1 in Fig. 7(b-1) correctly, but the detected human in Fig. 7(c-1) corresponded to human-1 in Fig. 7(b-1) incorrectly. However, false-correspondence in Fig. 7 (b-1) was corrected in next image (Fig. 7(c-2)). Next, the front face area was detected by face detection method using Haar-like features in Fig. 7(a-3) and it was labeled human-2. The humans in Fig. 7(b-3) and Fig. 7(c-3) were detected by information of human-2 in Fig. 7(a-3). Also, in case a front face was never detected by face detection method using Haar-like features in viewpoint (c), a face was detected by information of the other viewpoints. Then, detected human was tracked by the mean-shift tracker. Trasjectories of human-1 was represented as orange dots and trajectories of human-2 was represented as green dots in Fig. 8.

Experimental results showed the effectiveness of the proposed method.

6. CONCLUSION

In this paper, we proposed a multi-viewpoint human tracking method based on face detection using Haar-like features and mean-shift tracker. We confirmed effectiveness of the proposed method by experimental result. Also, we proposed a reduction method of false-

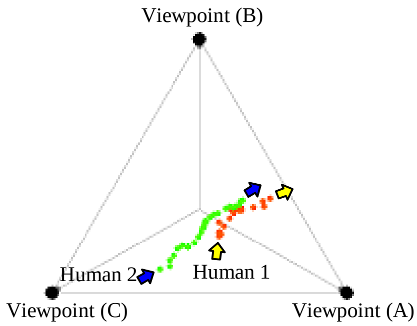


Fig. 8 Trajectories of humans

correspondence between viewpoints and confirmed effectiveness of the reduction method.

As a future work, we should remove assumption that a hair color is black.

REFERENCES

[1] 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, pp.511-518, (2001)

[2] 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)

[3] 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)

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

[5] 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)

[6] 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)

[7] Gray R. Bradski: "Real Time Face and Object Tracking as a Component of a Perceptual User Interface", Proceeding of the 4th IEEE Workshop on Applications of Computer Vision, pp.214-219, (1998)

[8] Gael Jaffre and Philippe Joly: "Costume: A New Feature for Automatic Video Content Indexing", Proceedings of RIAO2004, pp.314.325, (2004)

[9] 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)