Human Tracking Using Subtraction Stereo and Color Information

Yuma Hoshikawa
Chuo University / CREST, JST
hoshika@sensor.mech.chuo-u.ac.jp

Kenji Terabayashi
Chuo University / CREST, JST
terabayashi@mech.chuo-u.ac.jp

Kazunori Umeda
Chuo University / CREST, JST
umeda@mech.chuo-u.ac.jp

ABSTRACT

In this paper, we propose a method for human tracking using a stereo camera system called "Subtraction Stereo" and color information. The tracking system using the subtraction stereo, which focuses its stereo matching algorithm to regions obtained by background subtraction, is realized using Kalman filter. To make the tracking system more robust, the new method also uses color information as another distinctive information of person. The effectiveness of the proposed method is verified in the scene which is difficult to realize without color information.

Author Keywords

Human Tracking, Kalman Filter, Color Histogram, Bhattacharyya Distance, Sensor Fusion.

ACM Classification Keywords

Security

1. INTRODUCTION

Detection and tracking of persons using visual information are the important task for many applications such as visual surveillance, automotive safety and human-robot interface. To accomplish this task, a huge number of studies have been carried out until now [4]. Especially in recent years, many studies about three dimensional human tracking for surveillance using stereo vision have been reported [1, 2, 6]. This paper deals with tracking of multiple persons for a surveillance use.

Considering the surveillance use, persons often occlude each other in a scene. This has an inevitable problem for tracking. Zhao et al. [6] proposed a method to solve this problem by using several cameras and biometric sensors such as an iris identification system. Though this method works with very high success rate, it requires many devices and can be used only in a limited space. Bahadori et al. [1] proposed a robust method to track persons using distance information and color information. This method also works with high success rate, but it can only be used in limited range.

We proposed a stereo vision method called "subtraction stereo" [5] which realizes robust distance calculation by applying its stereo matching only to the extracted regions obtained by background subtraction. As this method focuses its distance calculation to foreground, it can detect persons from a disparity image in wide range with one stereo camera. We use this subtraction stereo for tracking.

For the tracking, we apply the three dimensional information obtained by subtraction stereo to Kalman filter. However, if we use only position information of person for tracking, the tracking can become unstable during occlusion. In order to reduce this problem, we propose a new method which uses color information to support tracking.

This paper is organized as follows. In section 2, we show the outline of the subtraction stereo. In section 3, we discuss the tracking method. In section 4, we present experimental results. And then, we conclude this paper in section 5.

2. SUBTRACTION STEREO

We show the basic algorithm of the subtraction stereo [5] in Figure 1. The subtraction stereo extracts objects in a scene by background subtraction method first, and then applies the stereo matching to the extracted regions. From the disparity image obtained by the subtraction stereo, actual heights and widths of the objects can be obtained. This size of the object can be used to identify whether extracted objects are persons or not. An example of output disparity image is shown in Figure 2.

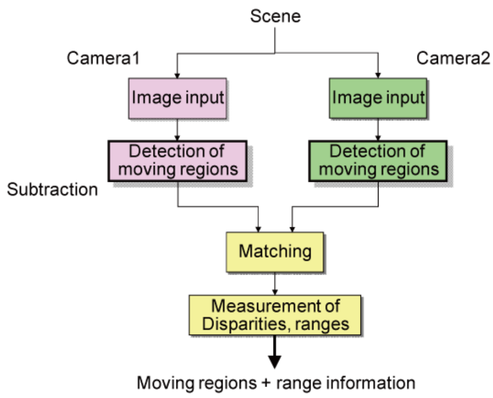
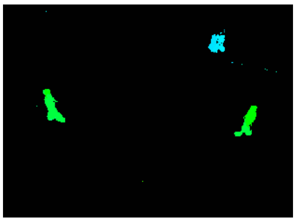


Figure 1. The basic algorithm of the subtraction stereo

(a) Input image



(b) Output disparity image

Figure 2. Disparity image obtained by subtraction stereo

3. HUMAN TRACKING

In this section, we explain the method to track multiple persons using Kalman filter and color information. Kalman filter is the method to estimate the state of a linear dynamic system from a series of noisy measurements by repeating observation and prediction alternately.

3.1 Tracking system using Kalman filter

In this paper, we apply a constant-velocity model for the state transition model of Kalman filter because the velocity of ambulation through frames can be considered as constant.

The state X of the Kalman filter is defined as:

$$\mathrm{X} = \begin{bmatrix} x & \dot{x} & y & \dot{y} & z & \dot{z} \end{bmatrix}^T \tag{1}$$

where (x, y, z) and $(\dot{x}, \dot{y}, \dot{z})$ are the world coordinate and velocity of person in the world coordinate system. The Kalman filter predicts the state at time i+1 from the state at i as

$$\mathrm{X}_{i+1} = \Phi \mathrm{X}_i + \Gamma_i \omega_i \tag{2}$$

where ω_i is the process noise and Φ is the state transition model matrix. Φ is given by

$$\Phi = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

The measurement Z for the Kalman filter is defined as

$$\mathrm{Z} = \begin{bmatrix} u & v & d \end{bmatrix}^T \tag{4}$$

where u and v are the image coordinate of person in an image, and d is the disparity. The relation between state X and measurement Z is represented as follows:

$$\mathrm{Z}_i = f(\mathrm{X}_i) + v_i \tag{5}$$

$$f(\mathrm{X}_i) = \begin{bmatrix} \dfrac{x_i \cdot f}{z_i} & \dfrac{y_i \cdot f}{z_i} & \dfrac{b \cdot f}{z_i} \end{bmatrix}^T \tag{6}$$

where f and b are the focal length and baseline length of the camera respectively, and v_i is the measurement noise. With all these variables, the Kalman filter predicts the three-dimensional coordinate of the person at each frame.

The tracking is done by associating the measured point with the predicted points. We define the set M which has elements m_j of the measured points, and the set K which has elements k_i of the predicted points. To associate the element m_j with the element k_i, the Euclidean distance $D_E(k_i, m_j)$ between all the measured points and predicted points are calculated. The measured points which satisfy both following conditions are associated with the predicted points.

- The Euclidean distance is less than the threshold D_{Th}.
- The Euclidean distance is the minimum.

In case of occlusion such that two persons walk toward each other and cross in front of the camera, only one measured point would be obtained even though there are two persons. Therefore, both two predicted points would be associated with the same measured point and the tracking becomes unstable. In this case, the Kalman filter do prediction without measured points, and after occlusion is over and each person is detected separately, each measured point is associated with predicted points again.

3.2 New tracking method using color information

The tracking method proposed in section 3.1 cannot cover all the situations which may happen in occlusion case. For example, if a person changes the direction during an occlusion, the predicted movement and actual movement of the person do not match and the tracking fails. To solve this problem, we propose a new method that uses color information to support the tracking.

For the color information, the Bhattacharyya distance [3], which represents the similarity between two normalized histograms, is used. In this research, the Bhattacharyya distance is calculated from color histograms which are made from hue values of the person. Figure 3 shows an example of the color histogram. We add this Bhattacharyya distance to the Euclidean distance with some weight. This new mixed distance is used to associate the measured points with the predicted points calculated by Kalman filter.

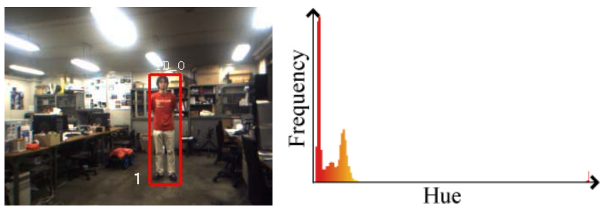
To use the color information for the tracking, the distinctive color information of each person must be obtained and color information model must be made. In this paper, the color histogram which is calculated as an average of first 10 frames is assumed to be the color information model of the person. The ID number of the color information model (the color histogram model) corresponds to the ID number of the Kalman filter. The Bhattacharyya distance between the model color histogram and the color histogram which is calculated from a current frame is used for tracking.

The Bhattacharyya distance D_B is given as

$$D_B = \sqrt{1 - \rho} \tag{7}$$

where ρ is the Bhattacharyya coefficient which is calculated from the following equation.

$$\rho = \sum_{u=1}^{m} \sqrt{p_u q_u} \tag{8}$$



(a) Scene **(b) Model color histogram**

Figure 3. An example of model color histogram

where p and q are the two normalized color histograms made from hue information, u is the hue number, and m is the number of the elements of hue. We define the new distance $D_{EB}(k_i, m_j)$ by combining the Euclidean distance and Bhattacharyya distance as

$$D_{EB}(k_i,m_j) = \alpha_i \cdot \frac{D_E(k_i,m_j)}{D_{Th}} + (1-\alpha_i)\cdot D_B(k_i,m_j) + \delta(\alpha_i)\cdot U(D_E(k_i,m_j) - D_{Th2}) \tag{9}$$

where $D_B(k_i, m_j)$ is the Bhattacharyya distance between ith color histogram model and the color histogram calculated from the jth measured person. The ID number of the color histogram model corresponds to the Kalman filter's ID number. We also define the set $M_{Th,i}$ having elements of the measured points whose Euclidean distance from the predicated point i is less than the threshold D_{Th}. The α_i in the equation (9) is defined as

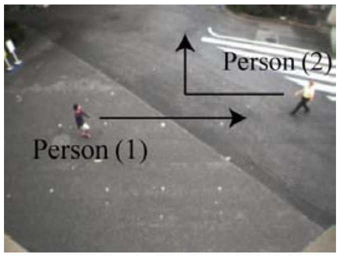
$$\alpha_i = \begin{cases} 0 & (n=0) \\ 1 & (n=1) \\ \min\limits_{s,t\in M_{Th,i}, s\neq t}\left(\dfrac{|D_E(k_i,m_s) - D_E(k_i,m_t)|}{D_{Th}}\right) & (n\geq 2) \end{cases} \tag{10}$$

where n is the number of elements in the set $M_{Th,i}$. α_i becomes small when several persons approach mutually and the tracking using only Euclidean distance becomes unstable. If there are measured points not so far from the predicted point but not belong to the set $M_{Th,i}$, and also not selected by any Kalman filter, α_i becomes 0 and only the Bhattacharyya distance is used to associate this measured point with the predicted point having ID number i. To decide whether the measured point is too far from the predicted point or not, another threshold D_{Th2} for the Euclidean distance is used. The new tracking method uses the new distance D_{EB} to associate the measured points with the predicted points.

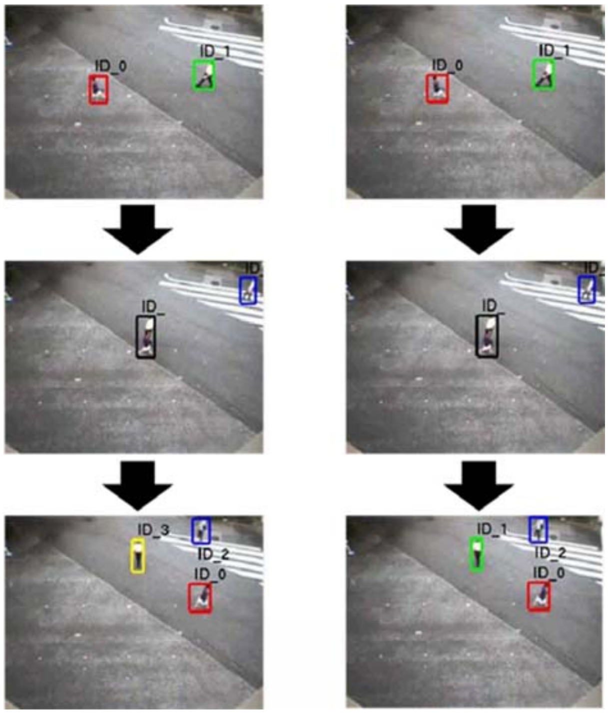
4. EXPERIMENTAL RESULTS

In this section, we show experimental results to evaluate the effectiveness of the new tracking method using color information. For the evaluation, we selected two scenes in which persons move unexpectedly during an occlusion. The evaluations are done by comparing the tracking results with the color information and without color information obtained from same offline data.

We show the experimental scene in Figure 4 (a) (scene1) and Figure 5 (a) (scene2). In scene1, two persons walk from the each side of the scene first, and when an occlusion occurs during intersection, one person changes the direction. In scene2, two persons walk from the each side of the scene first, and when an occlusion occurs during intersection, both two persons change the direction. For the both scenes, we regard the person walking from left end of the image as person (1) and the person walking from right end of the image as person (2).

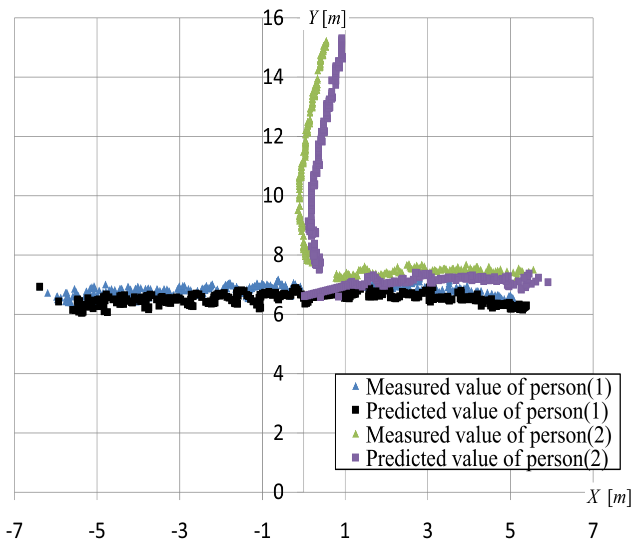


(a). Experimental scene (scene1)



Without color information **With color information**

(b). Experimental result of tracking



(c). Measured values of each person with color information

Figure 4. Experimental result for scene1

A stereo camera used for the experiments is Point Grey Research Bumblebee2 (color, f=3.8[mm]). The camera was set at the height of 8.3[m] with 50[^{o}] downward tilt. The process noise is set by 0.2[m/frame] for the velocity. The measurement noise is set to 0.3[pixel] for the image coordinate (u, v), and 0.04[pixel] for the disparity.

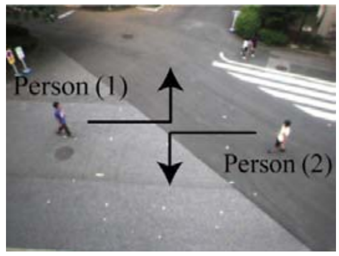
Figure 4 (b) shows the experimental result for the scene1, and Figure 5 (b) for the scene2. The colors of the frames around persons show the ID number of each person. Both results show that when the color information is used, each person is tracked well even if persons make movements which are difficult to be tracked without color information. And since the each person is tracked well, each measured point can be recognized in each frame, and trajectories for all the tracked persons can be obtained. Figure 4 (c) and Figure 5 (c) show the measured values and the predicted values of person1 and person2 obtained from scene1 and scene2 using the new tracking method. The graphs represent the top view of the experimental scenes. As the figures show, trajectories for each person are obtained well.

5. CONCLUSION

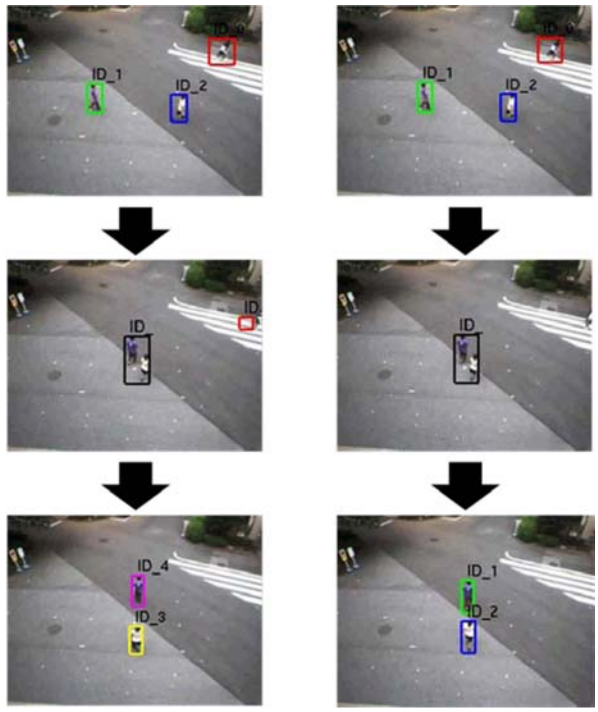
In this paper, we have discussed the subtraction stereo and proposed the new tracking method using subtraction stereo and color information. As the color information, Bhattacharyya distance is used. The effectiveness of the new tracking system was verified by the experiments for scenes which are difficult to realize without the color information.

REFERENCES

1. Bahadori, S., et al., “Real-time people localization and tracking through fixed stereo vision”, Applied Intelligence, Vol.26, No.2, pp.83-97, 2007.
2. Beymer, D. and Konolige, K., “Real-Time Tracking of Multiple People Using Continuous Detection”, *Proc. of the 7th Int. Conf. on Computer Vision Frame-rate Workshop*, 1999.
3. Bhattacharyya, A., “On a measure of divergence between two statistical populations defined by probability distributions”, Bull. Calcutta Math. Soc., Vol.35, pp.99–109, 1943.
4. Schiele, B., et al., “Visual People Detection-Different Models, Comparison and Discussion”, *Proc. of the IEEE ICRA2009 Workshop on People Detection and Tracking*, pp.1-8, 2009.
5. Umeda, K., et al., “Subtraction Stereo -A Stereo Camera System That Focuses On Moving Regions -”, *Proc. of SPIE-IS&T Electronic Imaging*, Vol.7239 Three-Dimensional Imaging Metrology, 723908, 2009.
6. Zhao, T., et al., “Toward a sentient environment: real-time wide area multiple human tracking with identities”, Machine Vision and Application, Vol.19, No.5, pp.301-314, 2008.

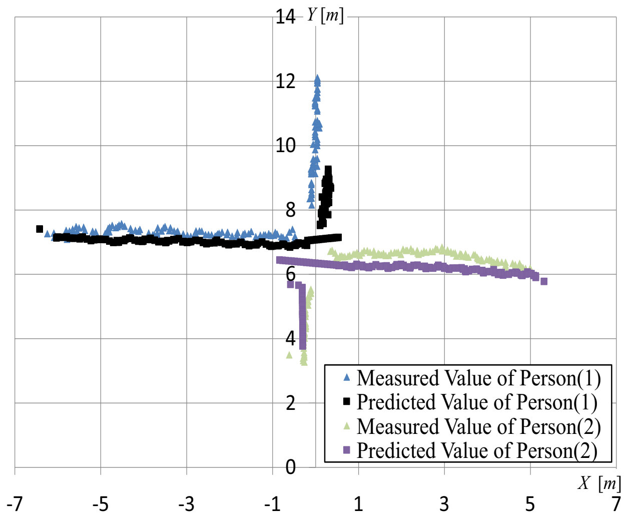


(a). Experimental scene (scene2)



Without color information **With color information**

(b). Experimental result of tracking



(c). Measured values of each persons with color information

Figure 5. Experimental result for scene2