Proc. Workshop of Machine Vision and its Applications, pp. 533-536, Tokyo, Japan, Nov. 2000.

Tracking Multiple Pedestrians in Crowd

Kyousuke Uchida *, Jun Miura, and Yoshiaki Shirai
Dept. of Computer-Controlled Mechanical Systems
Osaka University

Abstract

In the domain of city design, the demand for the development of the visual tracking method is increasing, which provides the pedestrian paths. This paper proposes the tracking method which reduces the ambiguity of matching. When features of a pedestrian are ambiguous in the original image, the proposed method generates plural candidates for the pedestrian path. In the following frames the proposed method verifies them to eliminate false ones. We experiment on real image sequences.

1 Introduction

In the domain of urban design, the designers are concerned that the flow of pedestrians is one of the most important keys to design the comfortable city[1]. For the analysis of the movements of crowds, they extract the pedestrian paths from videotapes which contain urban crowded scenes over an hour. In the scene there are always more than thirty pedestrians. Therefore the demand for the computer vision tracking method is increasing.

Conventional methods extract moving objects and track them [2]. There is a problem that overlapping objects are extracted as a region and tracked as a single person. We extract only isolated pedestrians and track them. By tracking them forward and backward, we can continue to track them even in a group.

By template matching [3], an operator often has to modify the position manually during the occlusion or when the color of the pedestrian is similar to that of adjacent pedestrians. The proposed method generates plural candidates for the pedestrian path and verifies them to eliminate false ones.

2 Extracting Pedestrians

We use the background subtraction method. First we model a background image from original images of sequence. Each pixel of the background is determined as the mode of R, G and B of the corresponding pixels of the images.

*Address: 2–1 Yamadaoka, Suita-shi, Osaka 565-0871 Japan. E-mail: uchida@cv.mech.eng.osaka-u.ac.jp

By the background subtraction method, a pedestrian often corresponds to several broken pieces. We merge the pieces to one region if a distance between the centroids of two pieces is smaller than a certain threshold.

In order to find an isolated pedestrian we select the object within a certain size as a pedestrian. In this setting of the camera, pedestrians range in size from 5 by 7 pixels to 8 by 20 pixels. (see Fig 2).

These extracted regions are used as the pedestrian template for tracking.

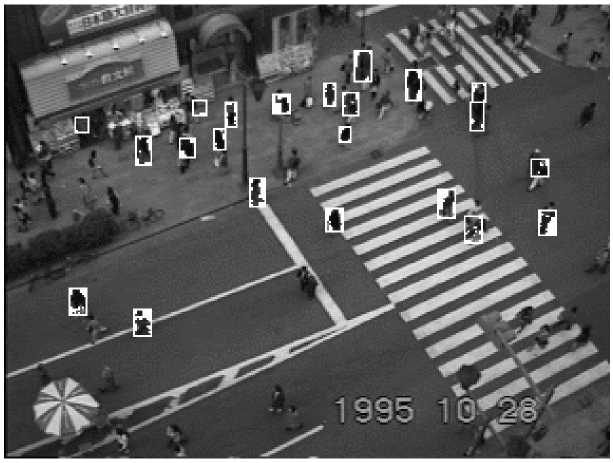


Figure 1: Extracting the image templates of pedestrians in the original image directly

3 Tracking Pedestrians

Because a pedestrian is often extracted on the way of walking, the tracking is executed not only forward but also backward.

3.1 Finding the Next Position of the Pedestrian

We limit the search area by predicting the movement of pedestrians. Since the pedestrian moves at an almost constant speed during a short period such as 1/15 second, the next position of the pedestrian can be predicted by linear extrapolation. We search only a neighboring area of the predicted position for a matching region. We thus avoid mismatch of pedestrians between frames.

For the evaluation of matching, we apply the SAD (Sum of Absolute Difference) to the matching score by equation1,2.

$$SAD = \sum_{j=1}^{n}\sum_{i=1}^{m}\left\{ \min_{-1\leq a,b\leq 1} AD_{(i,j),(i+a,j+b)} \right\} \quad (1)$$

$$AD_{(i,j),(k,l)} = max\{\mid R^T_{i,j} - R^I_{k,l}\mid, \mid G^T_{i,j} - G^I_{k,l}\mid, \mid B^T_{i,j} - B^I_{k,l}\mid\} \quad (2)$$

where m and n denotes the width and the height of the template, the superior letter T and I denotes the template and the original image, (i, j) and (k, l) denotes the position.

As the next position of the pedestrian, we select the position where the SAD score is local minimum and below a vairable threshold τ_t. If there are two and over such positions, we select the positions as the candidates.

The stability of the proposed method depends on the threshold τ_t. If the threshold τ_t is too high, too many next position candidates are found and selection true one becomes difficult. On the other hand, if τ_t is too low, the next position is not found. Furthermore, the suitable threshold τ_t per pixel (8bits) is from 8 to 20 depending upon the intensity of the pedestrian color (generally the high threshold needs to be set for the bright color pedestrian). Therefore we first set the threshold in proportion to the average of the intensity of the pedestrian. Since the suitable threshold approximates to SAD minimum score, in the following frame we updates the threshold by equation 3.

$$\tau_{t+1} = \begin{cases} \omega_1 \cdot \tau_t + \omega_2 \cdot SAD_m & (\tau_t > SAD_m) \\ \tau_t + c & (\tau_t \leq SAD_m) \end{cases} \quad (3)$$

where ω_1 and ω_2 denotes the weight coefficient for averaging, c denotes the constant coefficient. SAD_m is the minimum SAD score in the search area. If there are two candidates for the pedestrian path, SAD_m is the minimum SAD score in the two search area. Fig. 2 shows that the threshold approximates to the minimum SAD.

We, however, may not find next position of the pedestrian in the matching evaluation. Three cases are considered as sources.

1. The threshold τ_t is too small.
2. The pedestrian is occluded partially and we cannot get the good matching score.
3. The pedestrian is lost.

In these case, we continue searching the pedestrian in the following frames. About each source, we can find the pedestrian again when

1. The threshold τ_t is lifted up sufficiently.

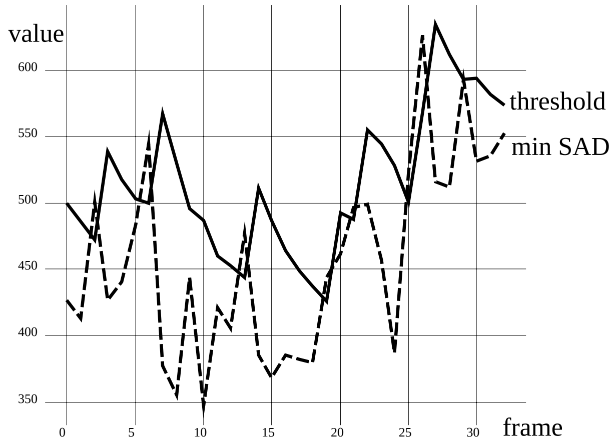


Figure 2: Minimum SAD and threshold τ_t

2. The pedestrian becomes isolated.

If we does not find the pedestrian again within the constant following frames, we admit losing the pedestrian (case 3.) and terminate tracking.

3.2 The Plural Candidates for the Pedestrian Path

On the other hand, we are often obliged to select plural positions of the pedestrian (see Fig.3(a)-(c)), because the features of the pedestrian is ambiguous in the original image when the pedestrian is occluded partially or the color of adjacent other pedestrians is similar to that of the pedestrian. To reduce such uncertainty, we keep plural candidates for the pedestrian path and verifies them to eliminate the false candidates in the following frames.

We provide two ways that can eliminate the false candidates. One is applied when the two and over paths merge in the same path. The one whose journey is shorter is selected and the others are rejected (see Fig.3(c)-(d)).

Another way is applied when tracking is terminated. Since it is difficult for the false path to get the SAD score below the threshold τ_t, the next position of the pedestrian cannot be found again within the constant frames and terminate tracking (see Fig.3(d)-(f)).

Finally, we regard the last single candidate as the reliable path of the pedestrian (see Fig.3(f)).

3.3 Eliminating false paths

After extracting the pedestrian paths, we verify them to eliminate false paths. If tracking is error like in Fig.4(b), two paths are extracted for the pedestrian B (see Fig.4(c)). For verification, We search the region where the template was extracted (shown by a square in Fig.4). If two paths go through the region, we consider that one is true and another is false. we assume that the straighter path is the true

Figure 3: Plural candidates for the pedestrian paths are generated and verified

one, since the pedestrians walk toward their destination directly. We estimate the straight degree of the path by equation4 where we divide its journey by its distance. We eliminate the less straight path whose score is bigger (see Fig.4(d)).

$$\begin{aligned} r &= \frac{l}{d} \\ &= \frac{\sum_{t_s+1}^{t_e} |\mathbf{x}(t) - \mathbf{x}(t-1)|}{|\mathbf{x}(t_e) - \mathbf{x}(t_s)|} \end{aligned} \quad (4)$$

where time t_s and t_e denotes the start and the end of tracking.

3.4 The Experimental Result

The method is tasted with real videotapes containing an urban crowded scene. Fig.5 shows the

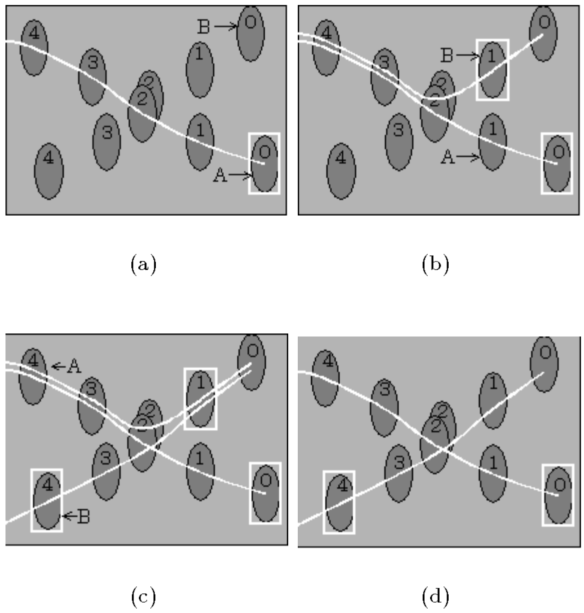


Figure 4: Detecting an error locus

pedestrian paths per second.

We notice that the pedestrian A makes his way curve to left in order to keep away from the pedestrian B and C.

We verify the experimental result by the human eyes. The Table 1 and Table 2 shows the efficiency of the proposed method.

Table 1: Experimental result of of tracking 1

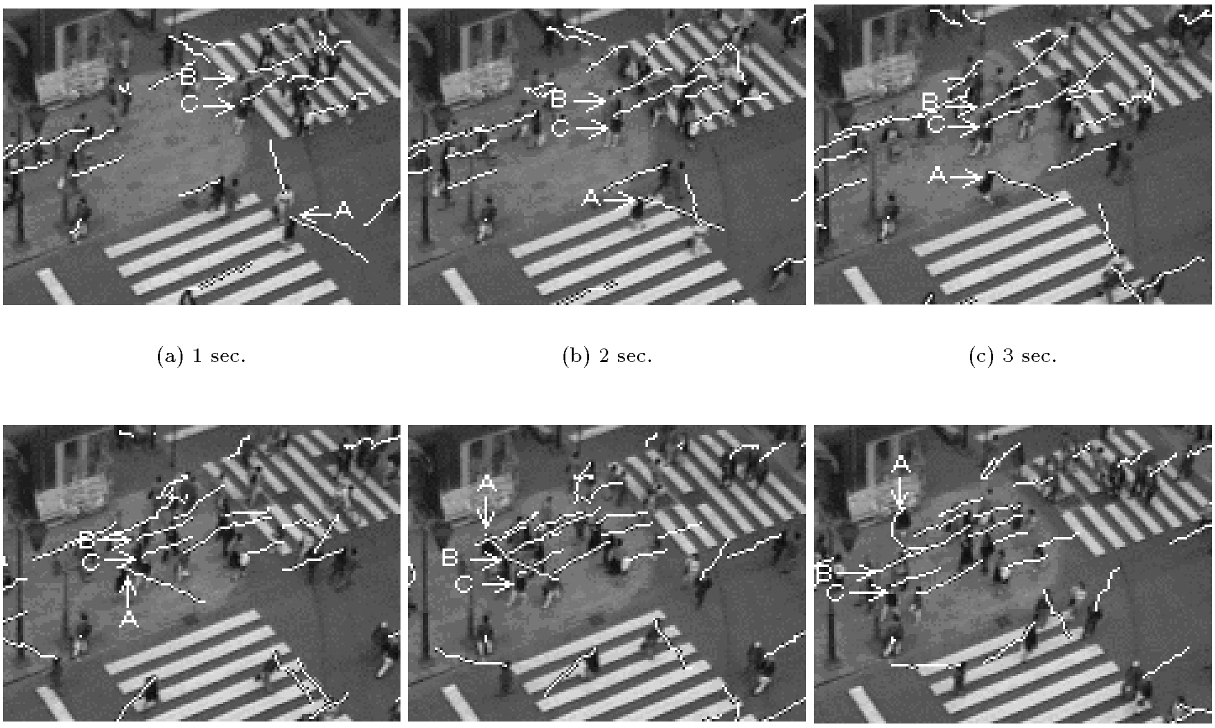
Pedestrians	Tracked	Missed
86	66 (77%)	20 (23%)

The proposed method provided the pedestrian paths about 77% (66/86) of all pedestrians in the sequence. The pedestrians whose color is similar to that of the background are apt to be missed since it is difficult to distinguish them from the background. Furthermore, the pedestrians overlapping through the image sequence are interpreted as a single pedestrian.

Table 2: Experimental result of tracking 2

Extracted paths	True	False
68	66 (97%)	2 (3%)

The 97% (66/68) of the extracted paths is reliable. When the color of adjacent other regions are similar to that of the pedestrian, tracking error is liable to occur. The proposed method was able to deal with such tracking problems, by generation and



(a) 1 sec. (b) 2 sec. (c) 3 sec.

(d) 4 sec. (e) 5 sec. (f) 6 sec.

Figure 5: The experimental result showing the pedestrian paths

verification the plural candidates for the pedestrian path.

4 Conclusion

In this paper, we described the method to track pedestrians in a complicated domain such an urban street. The proposed method was tasted with real videotapes and performed well tracking crowds of people in the street.

We are planning to taste the proposed method with more various scene and to analyze the movement of crowds based on the paths provided by the proposed method in future.

Acknowledgement

The authors would like to thank Dr. Kitai for providing us the image sequences.

References

[1] K. Kitai, "Space Activity - Quantitative analysis and evaluation with experiments on the street," Proc. S&T/SPIE Conference on Videometrics VI, pp. 268–275 (1999)

[2] M.Yachida, M.Asada and S.Tsuji, "Automatic Analysis of Moving Images," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.PAMI-3, No. 1, 12–20 (1981)

[3] S.S. Intille and J.W. Davis and A.F. Bobick, "Real-Time Closed-World Tracking," Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Press, pp. 697-703, (June 1997)