MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara - ken New Public Hall, Nara, Japan

8—39

Observation of Pedestrians with Multiple Views

Tokuyuki Mahara[1]
Department of Electronic and Mechanical Engineering
Tokyo University of Mercantile Marine

Tameharu Hasegawa[2]
Department of Computer Science
Chiba Institute of Technology

Heitoh Zen[3]
Institute of Media and Information Technology
Chiba University

Abstract

When pedestrians are observed using a view in the urban area, it is often difficult to track them because of the occlusions overlapped by roadside trees, traffic signs, advertising displays, street lamps, or another pedestrians. When multiple views are used for the observation of pedestrians, the state of the occlusion between views may be different. In this paper, we propose an algorithm to track pedestrians with multiple views taking an advantage of the differences of the state of their occlusion. In our proposal method, each pedestrian can be tracked even if a pedestrian is occluded at one view. Then, the location of an occluded pedestrian is estimated from the correspondence of pedestrians between views.

1 Introduction

In the field of urban planning engineering and civil engineering, pedestrians are observed with the aim of traffic facilitation. For the reason, it is necessary that each person should be tracked by the observation. In recent years, video sequences, which enable to observe pedestrians on space, are used in the observation[1]. However, if the video sequences are analyzed by persons, the analysis requires many persons and much hard labor. Thereby, an automation of the analysis from video sequences has been desired.

Many algorithms to track pedestrians with a single view have been proposed up to the present. In this case, it is often difficult to track each pedestrian, because a camera is set at the obscure place where a pedestrian is often overlapped by roadside trees, traffic signs, advertising displays, street lamps, or another pedestrians.

In this paper, we propose an algorithm to track each pedestrian with multiple views can observe the difference of the state of the occlusions between views as shown in Fig.1.In pre-processing, a homography matrix that a common plane between views is transformed from one view to another view is calculated. Every homography matrix to transform from one view to another view is required in main processing. In main processing, video sequences that synchronized between views are input. Then pedestrians are tracked at each view individually. After that, pedestrians between views are correlated using homography transformation. Then locations of occluded pedestrians are estimated by the correspondence of pedestrians between views.

2 Pre-processing of Proposal Method

Equation (1) is a homography matrix that transforms projectively from image coordinates (x, y) of one view onto image coordinates (u, v) of another view. Variables a, b, c, d, e, f, g and h are unknown. The variables are calculated from no less than four pairs of feature points[2].

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \cdots (1)$$

Pairs of the feature points between two views are acquired by hand as shown in Fig.3. Every homography matrix to transform from one view onto another view is acquired.

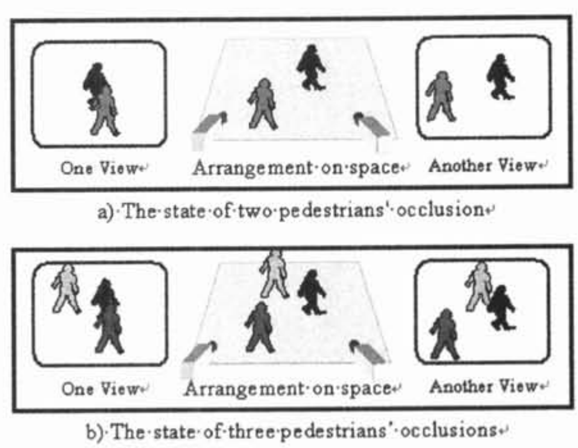


Fig.1: Comparison with occlusions between views

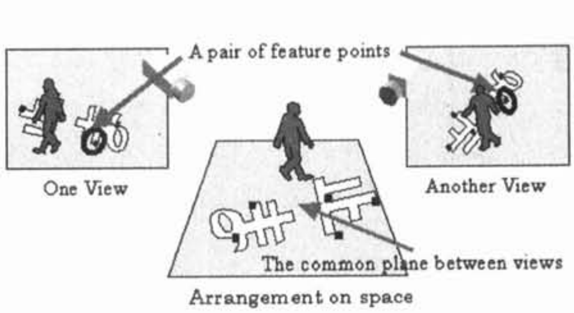


Fig.2: Pairs of feature points

---

[1] Address: 2-1-6 Etchu-jima, Koto-ku, Tokyo 135-8533 Japan. E-mail: tmaha@ipc.tosho-u.ac.jp

[2] Address: 2-17-1 Tsudanuma, Narashino, Chiba 275-0016 Japan. E-mail: hasegawa@cs.it-chiba.ac.jp

[3] Address: 1-33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522 Japan. E-mail: zen@imit.chiba-u.ac.jp

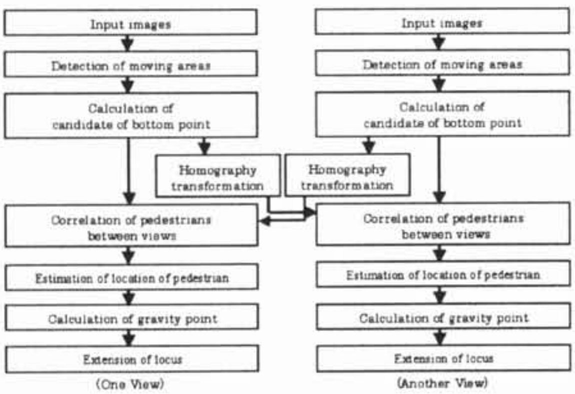


Fig.3: Flow diagram of proposal method

3 Main Processing of the Method

3.1 Input Images and Detection of Moving Regions

Overview of proposal method is shown in Fig.3. All video sequences that synchronized between views are input toward time series. Then, in order to detect moving areas, a current frame is subtracted from a previous frame. In a practical sense, in order to obtain moving regions enough to track pedestrians, the subtraction rate becomes longer than a video frame rate in Fig.4. Then the subtracted image is expanded and is binarized. We call this image as a subtraction image.

This method to detect moving region is effective for the image that a size of a pedestrian is smaller than a size of an input image.

3.2 Calculation of Candidate of Bottom Point

By means of method of Labeling, concluded components in subtraction image are labeled with identification number. Then, a circumscribed rectangle, an area, a gravity point and a candidate of a bottom point of the concluded components are calculated as shown in Fig.5. Next, if a concluded component is small enough to regard as noises, the concluded component is deleted from the image. We call the image as a label image.

Where a candidate of a bottom point is defined as the point that vertical line from the gravity point of the concluded component intersects with horizontal line from the lowest point of the concluded component.

3.3 Homography Transformation of candidates of bottom points

A common plane between two views is matched by homography transformation in three dimensions. In general a road surface as the common plane between two views is selected. If a candidate of a bottom point is on the common plane between two views and the candidate of bottom point is transformed from one view to another view by the homography transformation, the candidate of bottom point in one view is matched with the candidate of bottom point in another view as shown in Fig.6.

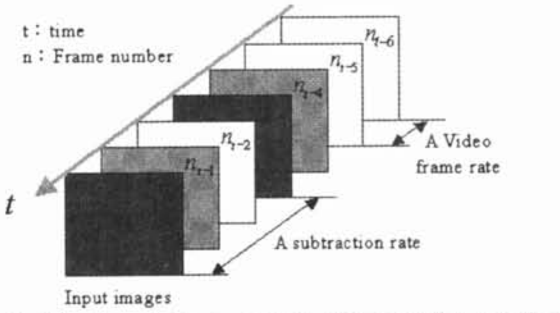


Fig.4: Subtraction both a current frame and a previous frame

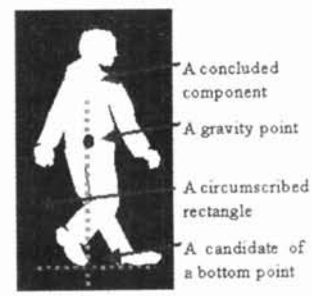


Fig.5: Moving region

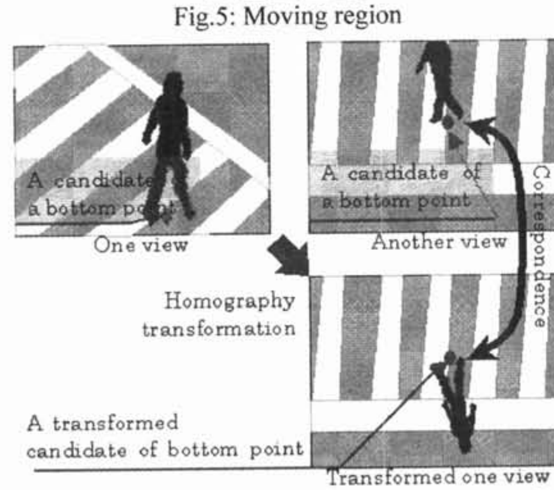


Fig.6: Homography transformation of a candidate of a bottom point

3.4 Correspondence of bottom point candidates between views

A candidate of a bottom point in one view is related with the nearest candidate of the bottom point that transformed from another view onto one view by the homography transformation as shown in Fig.7.

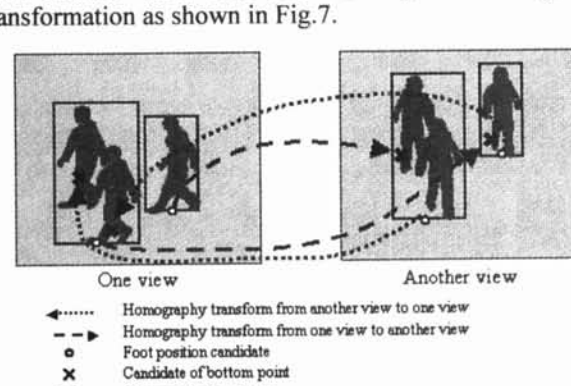


Fig.7: Correlation of candidate of bottom point between views

3.5 Estimation of the location of a Pedestrian

The candidates of the bottom point calculated in Section 3.2 are not always on the common plane between views, because the moving region of a pedestrian may be divided into some moving regions in subtraction images as shown in Fig.8-(a). Then, if the moving regions are detected as the one, the candidate of bottom point is only one and the location of bottom point is not always at the location of the bottom of a pedestrian as shown in Fig.8-(b). Thereby, in estimation of the location of a pedestrian, whether the candidate of bottom point in one view is described correctly as the location of a pedestrian in the view is determined. The estimation shows as follows.

(i) The candidate of bottom points in another view that related with watched candidate of bottom point in one view is checked up as shown in Fig.9.

(ii) A rectangle of proper size as a pedestrian is made. We call the rectangle as the rectangle to estimate the location of a pedestrian. Then the rectangles are set as the center of the bottom of the rectangle is located at the candidate of bottom position that transformed by the homography transformation from another view as shown in Fig. 10.

(iii) Every combination of the rectangle to estimate the location of a pedestrian is calculated as shown in Fig.11. Then, Equation (2) is calculated in the every combination.

(iv) The location of pedestrian is estimated, when the p is the largest in the combination as shown in Fig. 12.

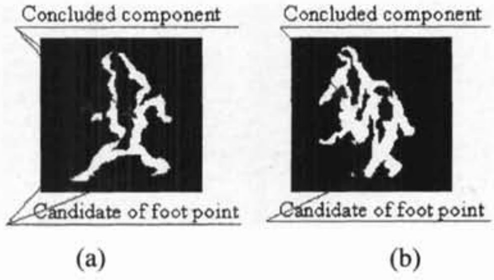
(v) If every concluded component in one view is estimated, the estimation of the location of a pedestrian is carried out in another view.

$$p = \frac{1}{2} \cdot \frac{S}{S_1} + \frac{1}{2} \cdot \frac{S}{S_2} \cdots (2)$$

Where, p is the value of evaluation that shows pedestrians' existing possibility; S is the area of concluded component that included in the rectangles to estimate the location of a pedestrian in a combination; S_1 is the area of concluded component in a combination. S_2 is the area of the rectangles to estimate the location of pedestrian in a combination.

4 Experiments

We observe pedestrians with two views, which slantwise look down at an observation area, in Matsudo city, Chiba, Japan, September 2001 as shown in Fig.13. It was occasionally cloudy.In order to synchronize view1 with view2 in the distant place, Horita system that involves application of GPS was used [3].



(a) (b)

Fig.8: Relation both moving areas and candidates of bottom points

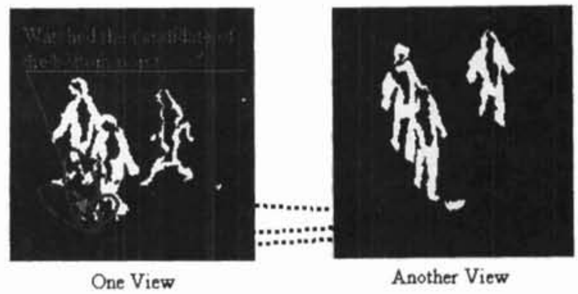


Fig.9: The candidate of bottom point related with watched the candidate of the bottom point

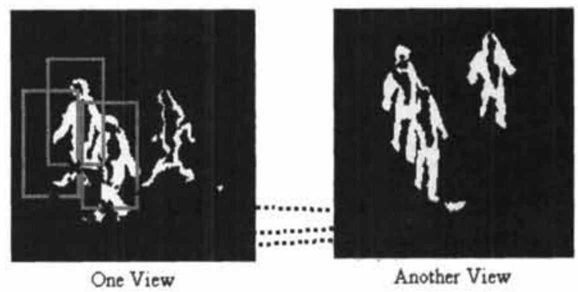


Fig.10: Setting of the rectangle to estimate the location of a pedestrian

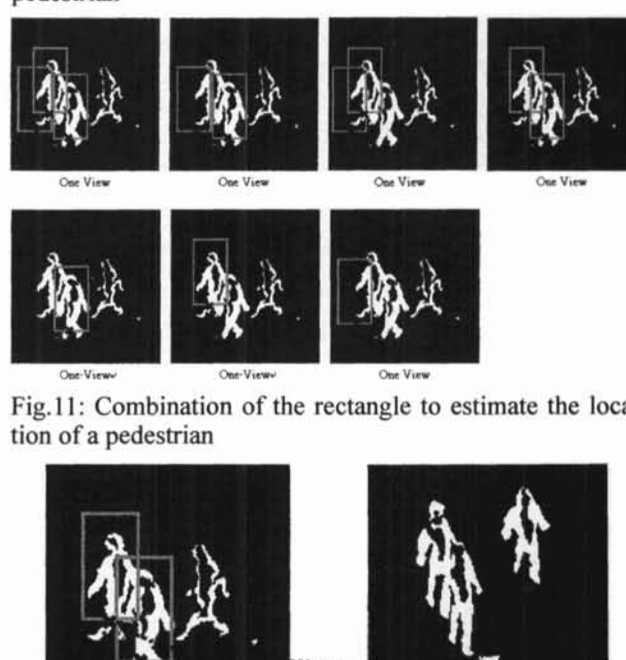


Fig.11: Combination of the rectangle to estimate the location of a pedestrian

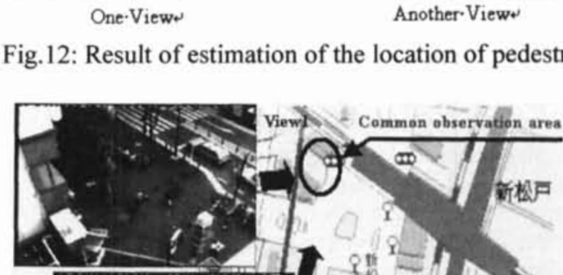


Fig.12: Result of estimation of the location of pedestrian

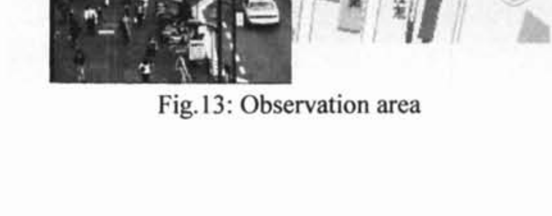


Fig.13: Observation area

4.1 The Experimental Results

The results of experiment are shown as follows. Input images are shown in Fig. 14. Then, Subtraction images were calculated from these input images as shown in Fig.15. Although many moving regions except moving objects were detected in Fig. 15, these in subtraction images were deleted by taking advantage of labeling method as shown in Fig.16. In label images in Fig.16, every concluded component with identification number shows with color to distinguish them. After that, the candidate of the bottom point and the circumscribed rectangle of every concluded component were obtained as shown in Fig.17. The rectangles to estimate the location of a pedestrian were set as shown in Fig. 18. The rectangles are colored to distinguish the correlation of the candidate of the bottom position. Result of the estimation of the locations of the pedestrians is shown in Fig.19. The results of the estimation of the location of pedestrian ware better as shown in Tabel.1 and Tabel.2.

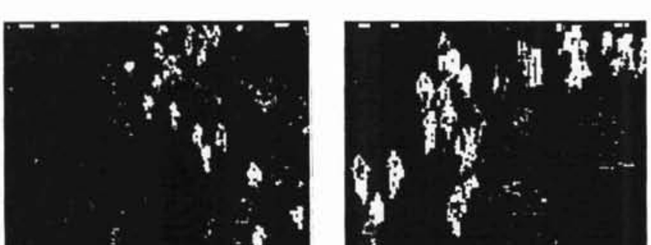
Tabel.1 ratio of correct estimation (View1)

Number of times to estimate	Correct estimation	False estimation
245	237(97%)	8(3%)

Tabel.2 ratio of correct estimation (View2)

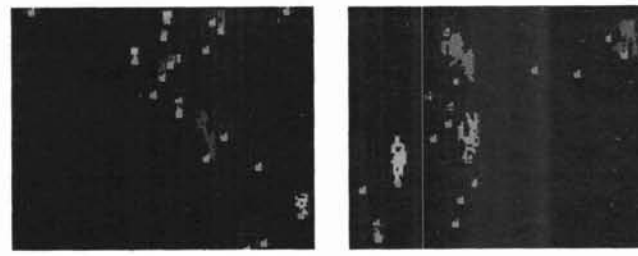
Number of times to estimate	Correct estimations	False estimations
343	335(98%)	8(2%)

a) View1 b) View2
Fig.14: Input images

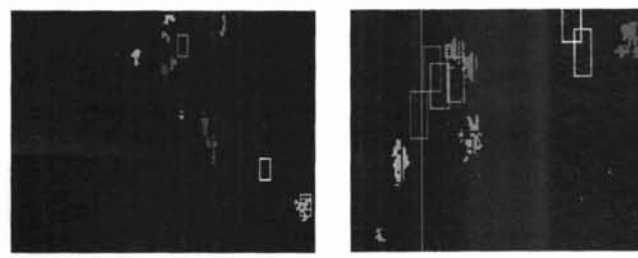


a) View1 b) View2
Fig.15: Subtraction images

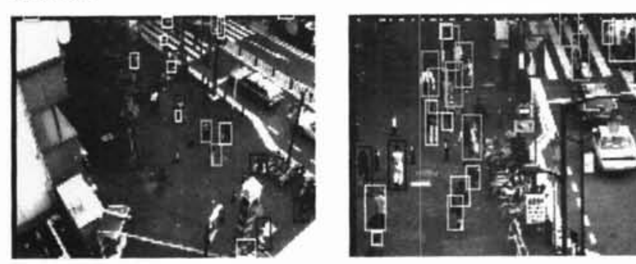
a) View1 b) View2
Fig.16: Label images



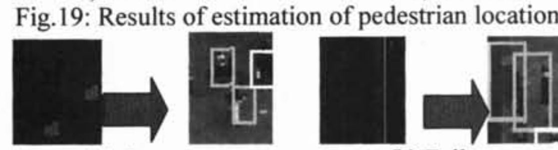
a) View1 b) View2
Fig.17: candidate of bottom point and circumscribed rectangle



a) View1 b) View2
Fig.18: Setting rectangles to estimate of location of a pedestrian



a) View1 b) View2
Fig.19: Results of estimation of pedestrian locations



(a) Success (b) Failure
Fig.20: Examples of estimation of the location of a pedestrian

5 Conclusion

In this paper, we proposed a method to track pedestrians in urban area and the estimation of the location of pedestrian. Our method was performed effectively. In our future work, we are planning to develop the tracking of pedestrians by expanding this estimation and information toward time series.

References

[1] Yutaka Suzuki, Naohiko Hibino, Yuichi Mouri, Tetsurou Hyodo: " Prospect on analysis of pedestrians' behavior with video sequence," the 21st corrected papers of Japan Society of Traffic Engineers, pp.153-156, 2001.

[2] N Ohta, K Kanatani: " Accuracy bounds and optimal computation of homography for image mosaicing applications," Proceedings of the 7th International Conference on Computer Vision, pp.73-78, 1999.

[3] http://horita.com