Detection of Change in the Number of Humans in a Monocular Image Sequence for Pedestrian Motion Tracking

Hidetaka Koseki
School of Engineering,
The University of Tokyo
Kashiwanoha 5-1-1
Kashiwa-shi, Chiba, JAPAN
koseki@race.u-tokyo.ac.jp

Morishita Soichiro
RACE,
The University of Tokyo
Kashiwanoha 5-1-1
Kashiwa-shi, Chiba, JAPAN
mori@race.u-tokyo.ac.jp

Hajime Asama
RACE,
The University of Tokyo
Kashiwanoha 5-1-1
Kashiwa-shi, Chiba, JAPAN
asama@race.u-tokyo.ac.jp

Abstract — Detection and analysis of trajectories are effective for providing services to pedestrians in public spaces. We have proposed a method to detect moving objects from a monocular image sequence with a normal mixture model. However, the method cannot accommodate a changing number of moving objects. As described herein, to overcome this shortcoming, we proposed two algorithms for counting number of moving objects automatically. They perform successive counting with detection of an increase or decrease of moving objects. We did an experiment using an image sequence from an actual environment to verify the availability of the proposed method.

Keywords: *monocular images, normal mixture model, multiple pedestrian detection, service provision*

I. Introduction

In a public space, analyzing tracks of pedestrians is an effective method for service providers to provide services for pedestrians[1]. Several recent studies have applied multiple sensors to track people. Ota et al.[2] proposed a method that uses optical ID sensors and floor sensors to obtain tracks. Murakita et al.[3] proposed a method that uses omnidirectional cameras, infrared sensors and floor pressure sensors to obtain tracks. Suzuki et al.[4] proposed a method that uses stereo vision cameras and a range finder to obtain tracks. These methods obtain accurate tracks. However, sensor installation costs are high. For that reason, installation of multiple sensor systems in existing spaces is difficult. Considering the installation cost, using a monocular camera for tracking is effective because these cameras are used on security cameras in business areas and crowded areas. For the reasons listed above, the purpose of this study is that of proposing a method to obtain tracks using a monocular camera. To obtain tracks of moving objects from video footage, two steps can be used.

1) Extract the area of a pedestrian from some video footage.
2) Obtain coordinate tracks of pedestrians' feet.

For the former step, Kamibata et al. proposed a method that is robust to changes of illumination conditions[5]. For the latter step, Morishita et al. proposed a method that identifies the foot position from the barycenter of the figure in the video footage[6]. Hirose et al. proposed a method to track multiple pedestrians using a normal mixture model[7]. This method approximates a pedestrian area as an ellipse. A barycenter and a variance–covariance matrix of a ellipse are computed for the pedestrian area. Finally, the method assumes the end of the long axis of the ellipse as the foot of the pedestrian. This method assumes that multiple pedestrians are a set of points of normal mixture distributions. Then the method obtains the barycenter of each pedestrian through normal mixture estimation. However, this method is applicable to video footage for which the number of pedestrians is given. In this paper, we propose a method to count the pedestrians using the information in the video footage.

II. Related Works

Related works of moving objects measurement are the following.

1) Template matching
2) Applying 3D model
3) Using spatiotemporal information

The following methods are examples of template matching. Curio et al.[8] proposed a method that applies a model to the gait of a pedestrian moving across a road. Hirono et al.[9] proposed a method that makes a model of moving video footage of the upper body. The method detects standing bodies by application of the model to standing bodies. Nakaue et al.[10] prepared a template made using statistical data of a pedestrian's upper body. Although the methods of Curio et al. are restricted to application of the video footage recorded with pedestrians from the lateral view. The method of Hirono et al. has the disadvantage that the detection rate changes according to the pedestrian position. The methods of Hirono et al. and Nakaue et al. only detect moving objects; they are not useful for tracking moving objects.

The following methods are examples of using 3D models. Kato et al.[11] proposed a method of a 3D model in which a pedestrian is approximated as an ellipsoid. Xiao-Wei et al. [12] made a 3D model in which the head of pedestrian is approximated as a sphere and a body of a pedestrian is approximated as a cylinder. The method described by Kato et al. [11] does not track many pedestrians. The method that Xiao-Wei proposed [12] is restricted to application to the environment because the video footage must be recorded directly above the pedestrian.

Kamijo et al.[13] proposed a method known as the spatio-temporal Markov Random Field Model. This method has robustness to occlusion, but its calculation cost is high.

III. Algorithm for Discriminating the Pedestrians and Discerning their Number

A. Method of Foot Detection and Tracking Algorithm

As described herein, we assume that all moving objects in a public space are pedestrians. First, we applied background subtraction[5] to the footage. Second, we compute a barycenter and a variance-covariance matrix from the extracted moving figure in the video footage. Then, we compute ellipse covering the figure using the barycenter and variance-covariance matrix. Finally, the end of long axis of the ellipse is defined as feet of pedestrians[6]. Barycenter and variance-covariance matrix of each pedestrian is computed by presuming a normal mixture model of EM algorithm if the number of pedestrian is given.

Then we describe a normal mixture model. Let n be the intended number of sample points obtained from the pixels of moving object. Then, $\boldsymbol{x}_n$ signifies the coordinate of the sample point, $\bar{\alpha}$ denotes the initial mixing degree, $\bar{\boldsymbol{\mu}}$ stands for the mean of the normal distribution, and $\bar{\boldsymbol{\Sigma}}$is a variance-covariance matrix. The EM algorithm with $\bar{\alpha}$, $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$ estimates posterior probability $\bar{Q}_n(k)$. In addition, k is the order of normal distributions.

As described in this paper, the mixing degree of each normal distribution is equal because occurrence probabilities of background subtraction pixels of each pedestrians are equal. When a normal mixture model estimation is applied to some video footage, the initial number of pedestrians in scene is restricted to one. The following parameters are initial values when estimation starts.

$$\bar{\alpha} = \frac{1}{k} \qquad (k = 1) \tag{1}$$

$$\bar{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \tag{2}$$

$$\bar{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \tag{3}$$

Then, maximum likelihood estimation using $\bar{Q}_n(k)$ and initial values derives feasible $\boldsymbol{\alpha}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Moreover, the following is the derived probability $p(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma_k})$.

$$p(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{\exp\left(-(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu}_k)/2\right)}{2\pi|\boldsymbol{\Sigma}_k|^{1/2}} \tag{4}$$

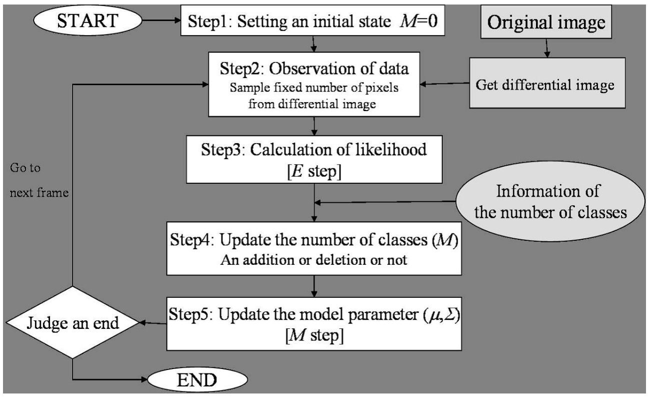
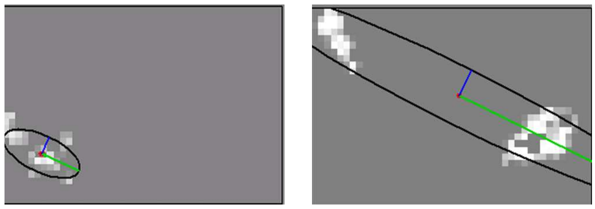


Fig. 1. Flow Chart for the Tracking Method



(a) Tracking Success (b) Tracking Failure

Fig. 2. Tracking moving object with $K = 1$

(4) means that sample marks of N units were generated with $p(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma_k})$ by the normal mixture distribution.

As described in this paper, K is the number of classes of a mixed normal distribution. M_k is the name of a class, and M represents the set of classes. The following shows the relation among these parameters.

$$M = \{M_k | k = 1, \ldots, K\} \tag{5}$$

The flow chart presented in Fig. 1 shows the method for tracking pedestrians. Fig. 2 shows the result of applying the mixture model estimation when K is 1. Fig. 2(a) shows the success of tracking when the number of pedestrians does not change. However, as the number of pedestrians increases, tracking fails (Fig. 2(b)). A cause of the tracking failure is that sample point coordinates $\boldsymbol{x}_n$ of two objects apply to one class.

B. Proposed Method

The analyses used in this paper subsume that the frame rate of the footage is sufficiently large and that the number of pedestrians changes in each frame in the footage is greater than one. The proposed method detects multiple pedestrians automatically. For each frame of the footage, the method corresponds to a change of the number of pedestrian and increase or decrease the number of classes K.

C. Pedestrian Number Determination

Fig. 2(b) shows that the discrepancy between the number of the class K and the number of pedestrians produces a tracking error. Morishita et al. obtained trained data to estimate one pedestrian's barycentric coordinate and a covariance matrix using the image of one pedestrian in advance. Then, Morishita et al. compared the trained data with other actually obtained data and determined the number of pedestrians in an elliptic region[15]. As described herein, S_{k1} and S_{k2} shows the long and minor axis of the ellipse produced from measured values. Here, S'_{k1} and S'_{k2} respectively signify long and minor axes of the ellipse, as inferred from estimated values. The following is the evaluation function used for this method of determination of the number of pedestrians.

$$\beta(M_k) = \begin{cases} S_{k1}S_{k2}/S'_{k1}S'_{k2} & (S_{k1}S_{k2} < S'_{k1}S'_{k2}) \\ S'_{k1}S'_{k2}/S_{k1}S_{k2} & (S'_{k1}S'_{k2} \leq S_{k1}S_{k2}) \end{cases} \quad (6)$$

To determine whether the number of pedestrians covered in the ellipse is one or not, we define $\beta(M_k)$ as the threshold of θ_1.

The trained data used for this study are the same as those described in a prior study [15].

1) Addition of a Class Using Likelihood of Sample Points: The following is the likelihood of sample points belonging to K-normal distributions.

$$lik(\boldsymbol{x}_n) = \sum_{k=1}^{K} p(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (7)$$

During the tracking function, the likelihood $lik(\boldsymbol{x}_n)$ of detected sample points that correspond to the noise or a pedestrian entering the image area is smaller than other $lik(\boldsymbol{x}_n)$. To discriminate noise and newcomers, this paper introduce threshold θ_2. When it comes to a set of class M, this paper determines O_M, which is a class of sample points as follows.

$$O_M = \{\boldsymbol{x}_n | lik(\boldsymbol{x}_n) < \theta_2\} \quad (8)$$

We decide a threshold θ_3 to assess a pedestrian entering the image. In other words, when $|O_M| > \theta_3$ is put into practice, O_M is the class of sample point of new pedestrian. Furthermore, we define additional class M_{K+1}.

2) Addition of Class Dividing the Class: In this chapter, we consider a case in which multiple pedestrians enter the image area with occlusion. In that case, if normal mixture estimation apply to multiple pedestrians area with $K = 1$, then the number of sample points of low likelihood are not always $|O_M| > \theta_3$. Therefore, in this case, the determination of increment the number of pedestrian using likelihood of sample points is unsuitable because the sample points distribution of multiple pedestrians with occlusion are similar to those of a single pedestrian.

To resolve the problem, we propose the following method. When a class M_k is $\beta(M_k) < \theta_1$, we divide the next frame sample points which belong to M_k by the minor axis of the ellipse. We sample points of M_k allotted to classes M'_k and M'_{k+1}. To allot sample points, we presume that the coordinate vector of sample points is $\boldsymbol{x} = (x_m,\ y_m)$, barycenter is $\boldsymbol{\mu} = (\mu_x,\ \mu_y)$ and that the gradient of the ellipse is ϕ. The following is an expression of the minor axis of the ellipse.

$$f(x) = \tan\phi(x - \mu_x) + \mu_y \quad (9)$$

A means to allot sample points using (9) is the following.

$$M_k = \begin{cases} M'_k & = \{\boldsymbol{x} | f(x_m) \leq m_y\} \\ M'_{k+1} & = \{\boldsymbol{x} | f(x_m) > m_y\} \end{cases} \quad (10)$$

If $\phi = 90$ [deg], then we allot sample points as shown below.

$$M_k = \begin{cases} M'_k & = \{\boldsymbol{x} | x_m \leq \mu_x\} \\ M'_{k+1} & = \{\boldsymbol{x} | x_m > \mu_x\} \end{cases} \quad (11)$$

D. Deletion of Class

When the number of pedestrians in the footage decreases and the number of classes K remain unchanged, tracking fails, as shown in Fig. 3. Fig. 3 shows that two classes apply to sample points of one pedestrian. $|O_M| \approx 0$ if no noise exists in the image of Fig. 3. A goodness of fit $\beta(M_k)$ does not detect the error like Fig. 3.

We propose a new set of classes N_l to detect a decrease in the number of pedestrians. We defined N_l as the set of L classes. Now, $L = K - 1$. The following is the relation between N_l of L classes and N of the set of the N_l.

$$N = \{N_l | l = 1, \ldots, L\} \quad (12)$$

When the normal mixture estimation identifies K pedestrians in the video footage, the estimation applies K pedestrians to M_k of K classes and N_l of L classes in parallel. In the case in which one is tracking K pedestrians, if tracking succeeds with class M_k, then goodness of fit $\beta(N_l)$ is $\beta(N_l) < \theta_1$. Then we consider the decrement of pedestrians in the video footage during the estimation tracks the K pedestrians. After one pedestrian leaves the image, the number of pedestrians in the image is L. Furthermore, the estimation successes with the set of classes N_l. Then, all of N_l consists $\beta(N_l) < \theta_1$. In this situation, we assess that the number of the class is greater than the number of pedestrians. We reduce class M_k.

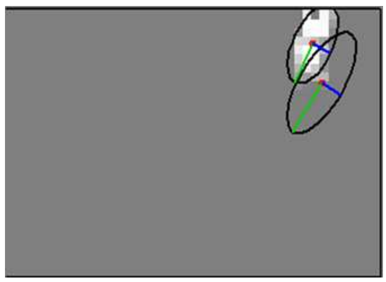


Fig. 3. Tracking failure caused by applying multiple classes to a pedestrian

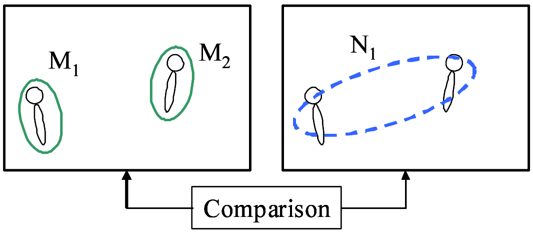


Fig. 4. Success of the tracking

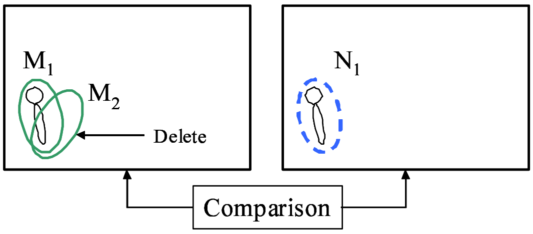


Fig. 5. Failure of tracking after decrement of the number of pedestrians

Fig. 4 and Fig. 5 show a case of class deletion with $K = 2$ and $L = 1$.

Fig. 4 presents the success of tracking using the set of classes M. When one pedestrian leaves the image area, tracking of class M_k fails, as shown in Fig. 5. In Fig. 4, the degree of fit is $\beta(N_1) > \theta_1$. In Fig. 5 however, $\beta(N_1) < \theta_1$ and we delete a class M_k. The mode of delete the class is that one of two classes applies to one pedestrian. Especially, the class which has the lowest likelihood of the class is deleted. The following is the expression can be used to assess likelihood.

$$\min \left\{ \sum_{k=1}^{K} p(\boldsymbol{x}_n; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \tag{13}$$

E. Procedure of the Proposed Method

We apply the algorithm when any pedestrian is visible in the video footage. When the algorithm detects a pedestrian at tth frames, we define the class as $K(t) = L(t) = 1$. For each frame, we apply the EM algorithm to the video footage. The method increments and decrements according to the following, which is the procedure of the method.

1) If the number of classes is $K(t) = L(t) = 1$, then
   - $\theta_3 < |O_M|$
     $\rightarrow K(t+1)$ is incremented by 1, using the method of III-C.1.
   - $\theta_1 > \beta(M_k)$
     $\rightarrow K(t+1)$ is incremented by 1, using the method of III-C.2
2) If the number of classes is $K(t) = L(t) + 1$, then
   - $\theta_3 < |O_M|$
     $\rightarrow K(t+1)$ and $L(t+1)$ are incremented by 1, using the method of III-C.1.
   - $\theta_1 > \beta(M_k)$
     $\rightarrow K(t+1)$ and $L(t+1)$ are incremented by 1, using the method of III-C.2.
   - $\theta_1 < \beta(N_l)$ and $L(t) > 1$
     $\rightarrow K(t+1)$ and $L(t+1)$ are decremented by 1.
   - $\theta_1 < \beta(N_l)$ and $L(t) = 1$
     $\rightarrow K(t+1)$ is decremented by 1.

When we increase or decrease the number of the classes, K and L must be greater than 0. Therefore the set of the classes is $L = 1$ when the algorithm is running with $K = 1$.

IV. Experimental Result

A. Experimental Conditions

To verify the validity of the proposed method, we experiment applying the method to the footage. The video images were recorded at the first floor entrance of Kashiwa Library at The University of Tokyo.

B. Experimental Methodology

Adaptive background estimation using M-estimation[5] i used to prepare a background image. Then we defined the experimental parameter empirically as I. Furthermore, we use goodness of fit β, which is the maximal value in the past five frames to avoid a false recognition of β because β of a class which stays on the verge of the image undergoes many changes. Moreover, after addition of a class using likelihood, β decreases for a few frames.

C. Far from Each Other

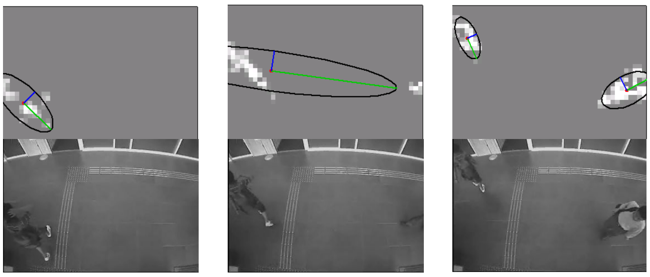
We apply video footage for which the number of pedestrians changes from one to two and then two to one.

1) Increment the Class: Fig. 6 shows the experimental result of the increment of pedestrians in the image area. Fig. 6(b) portrays an instantaneous frame in which a second pedestrian entered in the image area. At this frame, the result of the histogram of likelihood is Fig. 7. Regarding Fig. 7, sample points whose likelihood is less than $theta_2$ correspond to the second pedestrian area. Furthermore, those sample points form $\theta_3 < |O_M|$. Therefore, the class is incremented at the latter frame.

In the 14th frame, Fig. 6(b) shows that one class applies to the sample points of two pedestrians. Therefore tracking is not accurate at the frame, although tracking is successful after addition of the class after the 14th frame(Fig. 6(c)). Fig. 8 shows trajectories of the barycenter of two pedestrians. Figure in Fig. 8 portrays the frame number of the footage. A barycenter trajectory of a pedestrian who walks on the left side is distributed at the 14th frame because a normal mixture estimate computed a barycenter $\boldsymbol{\mu}_1$ using sample points of two pedestrians. Next we discuss a pedestrian who walks on the right side of the image.

TABLE I
EXPERIMENTAL PARAMETERS

Frame rate	15 [fps]
Number of sample points : n	230
threshold of goodness of fit $\beta(M_k)$: θ_1	0.25
threshold of likelihood $lik(\boldsymbol{x}_n)$: θ_2	1.0×10^{-5}
threshold of sample points class$\lvert O \rvert$: θ_3	23



(a) Tracking one pedestrian (8th frame)

(b) Pedestrian entering the image (14th frame)

(c) Increment the class (22nd frame)

Fig. 6. Experiment of increment the class using $\mathrm{lik}(\mathrm{x_n})$

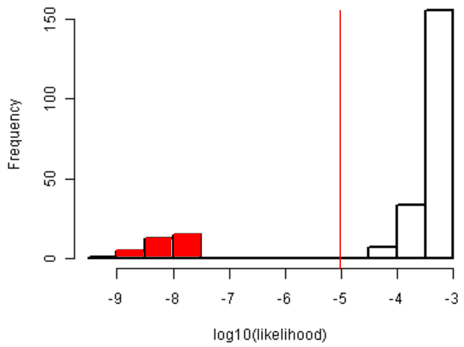


Fig. 7. Histogram of likelihood $\mathrm{lik}(\mathrm{x_n})$ (14th frame)

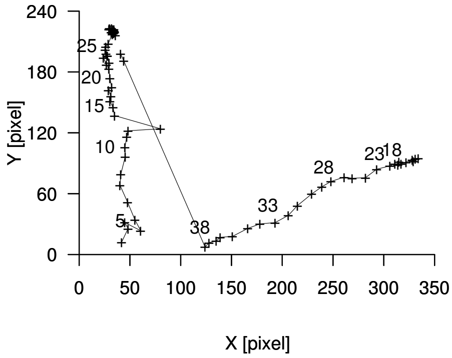
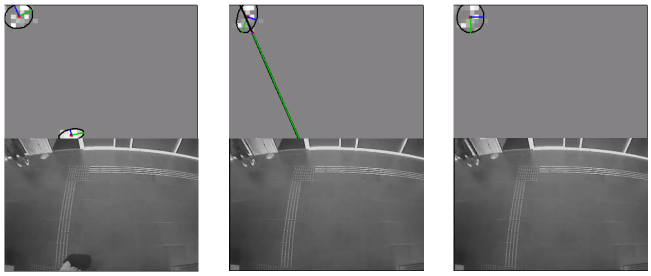


Fig. 8. Trajectories of the barycenter of two pedestrians



(a) Tracking two pedestrians (41st frame)

(b) A pedestrian exit (43rd frame)

(c) Delete the class (46th frame)

Fig. 9. Experiment deleting the class

2) Delete the Classes: Fig. 9 shows the experiment result in which one pedestrian exits the image area. After a pedestrian exits the image area, two classes apply to the one pedestrians sample points(Fig. 9(b)).

In this experiment, Fig. 10 portrays the goodness of fit β_{n1} of evaluation class N_1. At the 44th frame in Fig. 10, maximum β in the last five frames fulfills $\theta_1 < \beta(N_1)$. Therefore, the proposed method deletes the class. Before the method deletes a class, two classes belong to sample points of one pedestrian. Therefore, the trajectory of the class shows misalignment before deleting the class.

To obtain the proper barycenter trajectory, we must delete coordinates whose velocity vector changes enormously.

D. Footers Close to Each Other

We experimented to apply the method to video footage in which two pedestrians enter the image area alongside one another, and then part to the right and left. In Fig. 11(a), class M_1 belongs to sample points of the two pedestrian area. Therefore, tracking fails for the reason that the parameter satisfies goodness of fit is $\theta_1 < \beta(M_1)$, and the number of the low likelihood points is $\theta_3 > |O_M|$. However, the greater the distance of each pedestrian, the smaller $\beta(M_1)$ is. At $\beta(M_1) < \theta_1$, Fig. 11(c) shows the division of a class. Fig. 12 presents a transition of the goodness of fit $\beta(M_1)$. At the 53rd frame of Fig. 12, the maximum goodness of fit $\beta(M_1)$ in the past five frames is less than threshold θ_1. Therefore, a class is divided at the 54th frame. After the 54th frame, a goodness of fit $\beta(M_1)$ is larger than the threshold value. After dividing the class, class M_1 applies to a pedestrian of the right side and class M_2 applies to a pedestrian of the left side.

E. Evaluation of Experiments

We apply the proposed method to 6 footage including the situation described in Section IV-C or IV-D. The number of pedestrians in footage is two at most. We assumed it is successful in the case that all ellipses covering pedestrian's figure respectively. We evaluate every single frame in 6 footage. There are 428 frames in 6 footage, and 341 frames are successful. Success rate is 79.7%. We conclude that the success rate is sufficient by using only monocular camera and background subtraction.

V. Conclusion

As described herein, we proposed a method to infer the number of pedestrians in a monocular image sequence. Comparing the actual variance–covariance matrix and an estimated one, it was determined whether a moving object area is derived from a person or not. To detect increments of the number of pedestrians, it is assumed that sample points—the likelihood

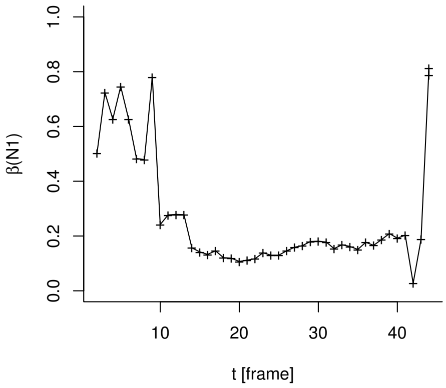


Fig. 10. Goodness of Fit β_{n1} of Evaluation Class N_1

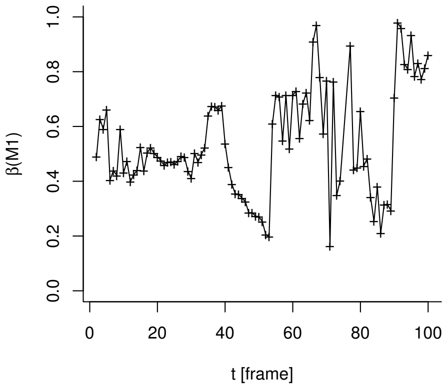
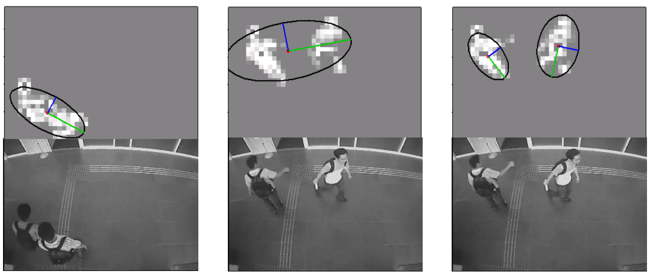


Fig. 12. Goodness of Fit $\beta(M_1)$ of Class M_1



(a) Defeat tracking (10th frame)

(b) $\beta_{M1} < \theta_3$ (53th frame)

(c) Division of a Class (54th frame)

Fig. 11. Experiment adding class by Division of a Class

of which are small—are derived from a person entering into the frame. Although this assumption does not hold when there is occlusion, we can divide the moving object area into two classes, comparing the size of the actual size of ellipse and the estimated one. To detect a decrement of the number of pedestrians, position identification of pedestrians is performed simultaneously under the assumption that the number of classes is fewer. Then the more appropriate one is adopted. Future work will apply this method to scenes with more than three pedestrians, and improve robustness to occlusion using a nonlinear model.

Acknowledgments

This work was part of the Intelligent Robot Technology Software Project supported by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

REFERENCES

[1] A. Nishimura et al.: "Estimation of Destination from Walking Patterns using Hidden Markov Model", JSME, ROBOMEC07, Vol.7, No.2, pp. 2P1-C10(1)–2P1-C10(3), 2007. in Japanese.

[2] S. Ota: "Multiple human tracking by integrating pixel-wise Optical ID sensors and floor sensors", Technical Report of IEICE, Vol.106, pp. 137–142, 2006

[3] T. Murakita et al.: "Dynamic Fusion of Visual Features Based on Multisensor Data", The 19th Annual Conference of JSAI, 2A3-05, 2005. In Japanese.

[4] T. Suzuki et al.: "Incorporating Environment Models for Improving Vision-Based Tracking of People", IEICE Transactions D-II Vol.J88-D-II, No.89, pp.1592–1600, 2005. In Japanese.

[5] J. Kamibata et al.: "Study on Shadow Detection using Correlation Analysis for Moving Object Extraction and its evaluation", JSME, ROBOMEC07, Vol.7, No.2, pp. 2P1-C09(1)–2P1-C09(4), 2007.

[6] S. Morishita et al.: "Study on Identification of Position of Foot from Barycentre of Figure in The Footage", SICE, SI2006, pp. 1380–1381, 2006. In Japanese

[7] T. Hirose, S. Morishita, H. Asama: "Foot position estimations for moving objects using a mixture model", IEEE International Conference on MFI, 344–349, 2008.

[8] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakasand, W. Seelen: "Walking Pedestrian Recognition", IEEE Transactions on Intelligent Transportation Systems, Vol.01, pp.155–163, 2000.

[9] A. Hirono et al.: "Measuring System for Distribution of People in Office", Matsushita Electric Works Technical Report 02855054, 2003/3, In Japanese

[10] N. Tomohiro et al.: "Real-Time human detection and counting system using human model and axial direction filtering", IEEJEISS, C, Vol.122-C,No.12, pp. 2011–2019 (2002/12) In Japanese

[11] H. Kato: "Real-time Human Tracking Using Ellipsoid Model", Transactions of Information Processing Society of Japan, 40(11), pp.4087–4096, 1999 In Japanese.

[12] X.-W. Xu et al.: "A rapid method for passing people counting in monocular video sequences", International Conference on Machine Learning and Cybernetics, Vol.3, 1657-1662, 2007.

[13] S. Kamijo et al.: "Vehicle Tracking in Low-angle and Front-View Images based on Spatio-Temporal Markov Random Field Model", 8th World Congress on ITS, Sydney Oct. 2001.

[14] A.P. Dempster et al.: "Maximum likelihood from incomplete data via the EM algorithm", Journal of Royal Statistical Society. Series B, Vol.39, No.1, pp.1–38, 1977.

[15] S. Morishita et al.: "Estimation of Distributed Parameter of Moving Object Region for a Person's Presence Determination", SI2007, 3L2-1, 2007 in Japanese.