A Method for Discriminating of Pedestrian based on Rhythm

Satoshi YASUTOMI and Hideo MORI

Department of Electrical Engineering and Computer Science
Yamanashi University
4-3-11 Takeda, Kofu, Yamanashi 400, Japan
E-mail: inu@kki.esi.yamanashi.ac.jp

Abstract

In this paper, a method for automatic pedestrian detection from monocular image sequence is described. It discriminates the pedestrians from the other moving objects based on the rhythmical motion of walking. When both feet of a pedestrian are on the ground in walking, their motions are found to have a relatively small changes of intensity in subtracted image. When one of the feet is moving forward, its motion is found to have a relatively large changes of intensity. Therefore, the periodical changing of image intensity caused by walking can be observed by applying Maximum Entropy Method. The state of a moving object, such as position and velocity, can be estimated with a kinematic model, a measurement model and a tracking filter. Corresponding to the *pace* and the 1/*stride* of walking, the rhythm is decomposed into two features; the temporal-frequency and the spatial-frequency. The model matching is performed based on these features. The method has the following advantages; (1) the rhythm is not affected so much by the kind of clothes, distance and weather; (2) the algorithm is simple enough to perform in real time. Experimental results in the asphalt paved road environment show the simplicity and the availability of this method.

1 Introduction

We consider the discriminating of pedestrians from the other moving objects from monocular image sequence. This discrimination is useful for many applications dealing with pedestrians, e.g., O'Rourke et al. [1] have analyzed human motion using constraint propagation and a detailed model of human body. A method for identification of human faces using eigenfaces has also presented by Turk et al. [2]. Such kind of methods are relatively high level identification methods and may require the rough information about the kind of the moving object in the earlier stage. However, this discrimination is difficult because of their various kinds of clothes and the changing of shapes. For these reasons, it is desirable to apply some method for such type of discrimination which employs some features independent of these influences.

There are some interesting reports related to such type of discrimination problem in the field of psychophisics. Johansson [3] has demonstrated the ability of subjects to perceive the body and the motion of a walker from Moving Light Display, and proposed a simple leg model. Kozlowski et al. [4] have applied the same method to demonstrate the ability of subjects to discriminate the sex of a walker as an abstract expression of motion. On the other hand, in the field of ethology, Tinbergen [5] has shown that the features of motion together with the other features are important elements for the animal behaviors. A butterfly called "*Eumenis semele*" reacts for the features of motion of its partner, and not for the features of shapes, colors and sizes. This behavior is an interesting idea for discriminating the pedestrians, which has the rhythmical motion of walking, from the other moving objects. Therefore, we make use of the rhythm of walking [6], and employ the temporal-frequency and the spatial-frequency as the features of rhythm for such type of discrimination. This method has the following advantages; (1) the rhythm is not affected so much by the kind of clothes, the distance between the object and the observer and weather; (2) the algorithm is simple enough to perform in real time.

The system of applied method is decomposed into three processes; (1) detection of moving object, (2) tracking, and (3) discrimination based on rhythm. There are various methods for the detection of moving objects such as subtraction of successive frames [7], subtraction from background image, optical flow [8], etc. We employ the subtraction of successive frames for the adaptability of slight change of environmental intensity and short processing time. Tracking methods can be categorized into two kinds of methods. One is based on the features of the object. Gilbert et al. [9] have presented an aircraft tracking system which is based on features obtained by two orthogonal projections. The other is based on the kinematic constraints of moving object. Legters et al. [10]

have applied a method which employed kinematic models and tracking filters to the simulated scenes. In most of the former methods, the objects are assumed to be rigid. Therefore, it is not suitable to track pedestrians as they have nonrigid body. So, we employ the latter method with the consideration of missing and outlying of the measurement.

In Section **2**, we describe the detection method of moving object and its position from the subtracted image. In Section **3**, we describe the tracking method which consists of a tracking filter and a validation investigator of measurement. In Section **4**, the methods for the detection of temporal-frequency and spatial-frequency and the model matching are presented. In Section **5**, experiments in the asphalt paved road environment show the ability to discriminate the pedestrians from the other moving objects. And finally, Section **6** gives some concluding remarks.

2 Detection of Moving Object

The video camera is fixed along the road so as to catch the moving object in the maximum field of view. The detection of moving object is performed in the following sequence.

1. The subtraction of successive frames and its absolution are performed.
2. To emphasize the moving object region and suppress the noises, Sobel vertical edge operator and the thresholding are performed in parallel. Then, the result of above operations are added.

This process is useful because the moving object regions include of many vertical edges more than the other regions.

3. The vertical and horizontal projections are performed, the bottom position and the width of the moving object region, that is, (u_p, v_p) and wid_{IMG}, are determined by slicing two projections.

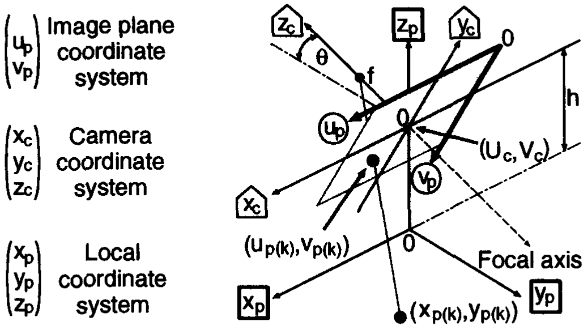
When the position of the object is predicted, all of the above processings are performed in a small detection window centered by the predicted position. Thus, the image processing cost and the influences caused by the other moving objects can be reduced.

3 Tracking

In the tracking system, it is desirable to continue the tracking whenever the measurement (u_p, v_p) is missed or outlied. Therefore, this requirement is implemented to continue the tracking at such conditions.

3.1 Kinematic and Measurement Models

The pedestrian is assumed to be a constant-velocity target, and represented by the model of discrete time state equation with sampling interval Δ(s). The video camera is assumed to be a pin-hole camera, and represented by the model of nonlinear function $\mathbf{h}_{(k)}$. The measurement model is depicted in **Figure 1**.



$\begin{pmatrix} u_p \\ v_p \end{pmatrix}$ Image plane coordinate system
$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}$ Camera coordinate system
$\begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix}$ Local coordinate system
z_c
z_p
y_c
0
θ
f
h
u_p
0
(U_c, V_c)
v_p
x_c
$(u_{p(k)}, v_{p(k)})$
0
Focal axis
x_p
$(x_{p(k)}, y_{p(k)})$
y_p

Figure 1: Measurement model and coordinate systems.

The state equation and the measurement equation are experssed as

$$\mathbf{X}_{(k+1)} = \mathbf{F}\mathbf{X}_{(k)} + \mathbf{w}_{(k)} \tag{1}$$

$$\mathbf{Y}_{(k)} = \mathbf{h}_{(k)}(\mathbf{X}_{(k)}) + \mathbf{z}_{(k)} \tag{2}$$

where

$$\mathbf{X}_{(k)} = \begin{bmatrix} x_{p(k)} \\ x_{v(k)} \\ y_{p(k)} \\ y_{v(k)} \end{bmatrix}, \mathbf{F} = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{w}_{(k)} = \begin{bmatrix} 0 \\ w_{1(k)} \\ 0 \\ w_{2(k)} \end{bmatrix}, \mathbf{Y}_{(k)} = \begin{bmatrix} u_{p(k)} \\ v_{p(k)} \end{bmatrix}, \mathbf{z}_{(k)} = \begin{bmatrix} z_{1(k)} \\ z_{2(k)} \end{bmatrix}.$$

$\mathbf{X}_{(k)}$ is the state vector consists of position (x_p, y_p) and velocity (x_v, y_v). $\mathbf{Y}_{(k)}$ is the measurement vector consists of position (u_p, v_p) on the image plane. $\mathbf{w}_{(k)}$ and $\mathbf{z}_{(k)}$ correspond to the system noise and the measurement noise which are sequence of zero-mean, white and Gaussian with covariance **Q** and **R**. **Q** and **R** are determined by experiments. $\mathbf{h}_{(k)}$ is the nonlinear function of $\mathbf{X}_{(k)}$ which consists of translations and perspective transformation expressed as

$$\mathbf{h}_{(k)}(\mathbf{X}_{(k)}) = \begin{bmatrix} U_c + \frac{a\ f\ x_{p(k)}}{f - y_{p(k)} \sin(\theta - \pi/2) + h \cos(\theta - \pi/2)} \\ \\ V_c - \frac{b\ f\ (y_{p(k)} \cos(\theta - \pi/2) + h \sin(\theta - \pi/2))}{f - y_{p(k)} \sin(\theta - \pi/2) + h \cos(\theta - \pi/2)} \end{bmatrix} \tag{3}$$

where f, h and θ correspond to focal length, height and depression angle of camera. (U_c, V_c) is the center position of image plane. a and b are the coefficients of aspect ratio.

We employ Extend Kalman Filter (EKF) for state estimation and prediction as follows:

$$\hat{\mathbf{X}}_{(k|k)} = \hat{\mathbf{X}}_{(k|k-1)} + \mathbf{K}_{(k)}\left[\mathbf{Y}_{(k)} - \mathbf{h}_{(k)}(\hat{\mathbf{X}}_{(k|k-1)})\right] \quad (4)$$

$$\hat{\mathbf{X}}_{(k+1|k)} = \mathbf{F}\hat{\mathbf{X}}_{(k|k)} \quad (5)$$

$$\mathbf{K}_{(k)} = \mathbf{P}_{(k|k-1)}\mathbf{H}^T_{(k)}\left[\mathbf{H}_{(k)}\mathbf{P}_{(k|k-1)}\mathbf{H}^T_{(k)} + \mathbf{R}\right]^{-1} \quad (6)$$

$$\mathbf{P}_{(k|k)} = \mathbf{P}_{(k|k-1)} - \mathbf{K}_{(k)}\mathbf{H}_{(k)}\mathbf{P}_{(k|k-1)} \quad (7)$$

$$\mathbf{P}_{(k+1|k)} = \mathbf{F}\mathbf{P}_{(k|k)}\mathbf{F}^T + \mathbf{Q} \quad (8)$$

where $\mathbf{H}_{(k)}$ is the Jacobian matrix expressed as

$$\mathbf{H}_{(k)} = \left(\frac{\partial \mathbf{h}_{(k)}}{\partial \mathbf{X}_{(k)}}\right)_{\mathbf{X}_{(k)}=\hat{\mathbf{X}}_{(k|k-1)}} \quad (9)$$

The state estimate $\hat{\mathbf{X}}_{(k|k)}$ is obtained each time by the incremental procedure of EKF. The predicted measurement vector is used for centering the detection window as mentioned in Section **2**.

3.2 Validation of Measurement

The occlusion by another moving object and the outlier of measurement by noises make the interference in the measurement and cause errors on the state estimation. Moreover, it will be ended the tracking in failure. For this reason, we confirm the validation of the measurement to determine whether the measurement should be used for the state estimation or not.

Occlusion When the image of pedestrian is covered by another moving object, occlusion may arise. In this case, the measured width of the object may show unsuitable value to the pedestrian width. So, the width of the object is checked for the detection of occlusion. The pedestrian width model is assumed to have a normal distribution $N(\overline{wid}, \sigma^2_{wid})$. The real width of the object wid is determined by applying inverse perspective transformation to wid_{IMG} at (u_p, v_p) , and checked by the following condition.

$$\frac{(wid - \overline{wid})^2}{\sigma^2_{wid}} \le k^2_{\sigma 1},\ (k_{\sigma 1} = const) \quad (10)$$

When this condition is satisfied, the object matches to the pedestrian width model, and the occlusion does not take place.

Outlier When the noises make outlier in measurement of position, the estimate may be incorrect for the valid estimation. A validation-gate is used to check the validation of the measurement. We employ the following condition as the validation-gate.

$$\left[\hat{\mathbf{X}}_{(k|k-1)} - \tilde{\mathbf{X}}\right]^T \mathbf{P}^{-1}_{(k|k-1)}\left[\hat{\mathbf{X}}_{(k|k-1)} - \tilde{\mathbf{X}}\right] \le k^2_{\sigma 2},$$

$$(k_{\sigma 2} = const) \quad (11)$$

where $\tilde{\mathbf{X}}$ is a temporal-estimate obtained by **(4)** with the measurement $\tilde{\mathbf{Y}}$ at (k). $\mathbf{P}_{(k|k-1)}$ is a state prediction covariance obtained by **(8)**. When this condition is satisfied, the measurement $\tilde{\mathbf{Y}}$ will be judged not to be outlier.

As mentioned above, when the measurement satisfies two conditions **(10)(11)**, the measurement is judged valid. The context of data flow diagram is depicted in **Figure 2**. When the validation of measurement is obtained, the state estimation is performed with the measurement $\tilde{\mathbf{Y}}$, that is, the gate switch is turned to the left side. When the validation of the measurement is not obtained, the state estimation is performed with the predicted measurement vector $\hat{\mathbf{Y}}_{(k|k-1)}$, that is, the gate switch is turned to the right side.

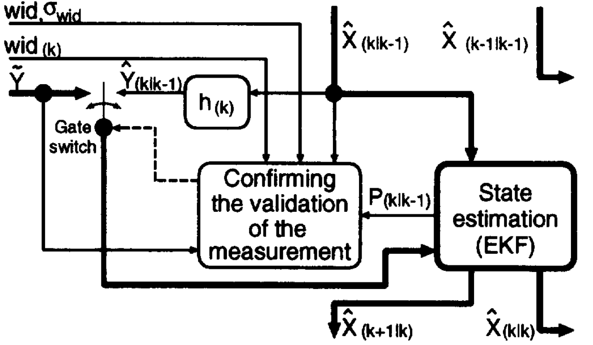


Figure 2: Confirming the validation of measurement and state estimation.

The state transition diagram of tracking process is depicted in **Figure 3**. "NONE" indicates that there is no moving object. When moving object is detected, the state is changed to "DETECT." Here, the detected position is compared with the detection range which is defined as the distance between the object and the observer. When the position is out of the range, the state returns to "NONE." In the other case, the initialization for the tracking process is performed, and the state is changed to "OBSERVE." "OBSERVE" indicates that the measurement is obtained with its validation and the state estimation is done by turning the gate switch to the left side as shown in **Figure 2**. When the valid measurement is not obtained, the state is changed to "PREDICT" and the predicted measurement vector is obtained by turning the gate switch to the right side. The valid measurement changes the state to "OBSERVE" again. When the estimated position is out of the detection range or a constant time is passed at "PREDICT," the state is changed to "NONE" which means the missing of moving object.

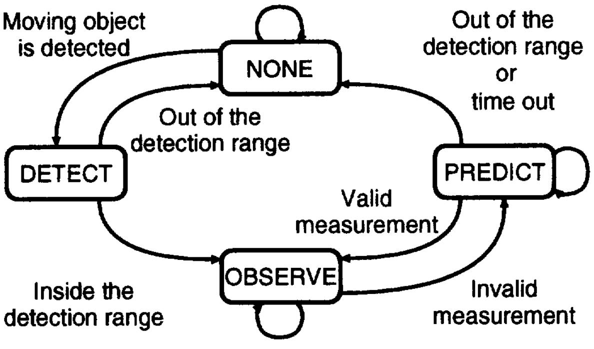

Moving object is detected
NONE
Out of the detection range or time out
Out of the detection range
DETECT
PREDICT
Valid measurement
OBSERVE
Inside the detection range
Invalid measurement

Figure 3: State transition diagram of tracking process.

4 Discrimination based on Rhythm

The walking is divided into two stages; first, both feet are on the ground, and second, one of the feet is moving forward. In the former, the motion of feet is relatively slow. In the latter, the motion of a foot is relatively fast. This phenomenon is found as the periodical changing of intensity on the subtracted image. When the temporal-frequency and the travelled distance of the moving object are estimated, the estimation of the spatial-frequency will be possible. Here, the detection and the model matching for the rhythm is described.

4.1 Detection of Rhythm

At the first stage, the process of moving object detection performs the subtraction of successive frames. We set up window "B" at the bottom of moving object region in the subtracted image as shown in **Figure 4**. The size of the window is W (the average width of human in the image) in width and $W/5$ in height. In this window, the summation of intensity is obtained as $I(k)$. The intensity summation in the window "B" is performed during the tracking. Sampling of $I(k)(k = 0, \cdots, N-1)$ takes $N \times \Delta$ seconds. The power spectrum $P(f)$ is obtained by applying Maximum Entropy Method (MEM) to the data $I(k)$. The order of Auto Regressive (AR) model for MEM is determined by experiments. Thus, we can obtain the principal temporal-frequency F. The spatial-period T is obtained as

$$T = \frac{D}{F \cdot N \cdot \Delta} \qquad (12)$$

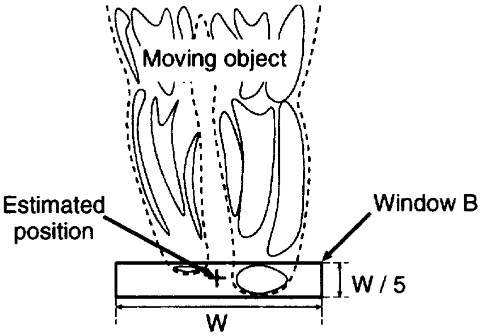


Figure 4: Window for detection of intensity changing.

where D denotes the travelled distance during the tracking in N cycles. Eventually, the temporal-frequency F and the spatial-frequency $1/T$ are obtained as the features of the rhythm.

4.2 Model Matching

If the moving object is a pedestrian, the temporal-frequency and the spatial-period represent the pace and the stride respectively. The model of typical walking is assumed that the pace and the stride have a normal distribution $N(\overline{F}, \sigma_F^2)$ and $N(\overline{T}, \sigma_T^2)$ respectively. The result of matching: Pedestrian or Non-pedestrian is judged by the following sequence.

1. The peak frequencies $\{F_l\} : (l = 1, \cdots, N_p)$ which have the power $\{P_l\}$ greater than P_{thr} are extracted. The summation of the powers is obtained as

$$S = \sum_{l=1}^{N_p} P_l. \qquad (13)$$

When N_p is 0, the result of matching is judged as Non-pedestrian.

2. The spatial-periods $\{T_l\}$ are obtained by applying **(12)** to the temporal-frequencies $\{F_l\}$ and the travelled distance D. Here, let us call $\{P_l\}$, $\{F_l\}$ and $\{T_l\}$ as "peak-set."

3. The peak-sets are tested whether these satisfy the following conditions or not.

$$(F_l - \overline{F})^2/\sigma_F^2 \le k_{\sigma 3}^2, \ (T_l - \overline{T})^2/\sigma_T^2 \le k_{\sigma 3}^2,$$

$$(k_{\sigma 3} = const) \qquad (14)$$

The peak-sets which satisfy the above conditions are renewed as $\{P_m\}$, $\{F_m\}$ and $\{T_m\}$: $(m = 1, \cdots, N_{pm})$. When N_{pm} is 0, that is, any peak-set does not satisfy the condition **(14)**, the result of matching is judged as Non-pedestrian.

4. The evaluation value $\{R_m\}$ are calculated which denotes the ratio of the peak and the other peak powers.

$$R_m = \frac{P_m}{S}, \; (m = 1, \cdots, N_{pm}) \tag{15}$$

5. The peak-set which has the greatest R_m is extracted. The result of matching is judged by the following conditions.

$$\begin{array}{llll} R_{thr} & < R_m \leq & 1.0 & \text{Pedestrian} \\ 0.0 & \leq R_m \leq & R_{thr} & \text{Non-pedestrian} \end{array}$$

When the object is judged as Pedestrian, the model matching is inactivated, and only the tracking process will be continued.

5 Experimental Results

In the experiments, the image sequences were taken in the asphalt paved road environment in the university campus. The video camera is fixed at 1.0(m) in height, 15.0(°) in depression angle. NTSC signal of the video camera "SONY CCD-TR1000" is decoded into the RGB components. In this method, G-component is only used which consists of about 60% luminous intensity. We use an image processing system "ANDROX ICS-400XM9" which consists of a master CPU (68030, OS-9) and 4-DSPs with frame memories of 512 × 512(pixel) × 8(bit). The range of moving object detection is set up to 3.0 ~ 20.0(m) in distance. The moving object detection and the tracking are performed in every 66(ms), namely 15(Hz). As mentioned in Section **3** and **4**, the model of typical walking and the condition parameters are set up as shown in **Table 1**.

Table 1: Parameters for the experiments.

Pace (Hz)	$\overline{F} = 2.050$, $\sigma_F = 0.183$
Stride (m)	$\overline{T} = 0.720$, $\sigma_T = 0.090$
Num. of intensity data	$N = 64$
AR model order	$M = 15$
Threshold of power	$P_{thr} = 10.0$
Threshold of evaluation	$R_{thr} = 0.1$
Approval threshold	$k_{\sigma 1} = k_{\sigma 2} = k_{\sigma 3} = 3.0$

Figure 5 shows two examples of the image sequence which were manually thinned out at interval of 1 ~ 5(s). These include 9 objects. The experimental results are shown in **Figure 6,7** with the notations of the result of discrimination, locus, F (pace), T (stride), R, velocity, intensity changing and power spectrum. Objects (1) ~ (6) are pedestrians with various types of clothes. Objects (7) and (8) are bicycles, and object (9) is a dog. Objects (7) ~ (9) were judged as Non-pedestrian as there is no peak-set which matches with the model of typical walking.

We assume that the pace of pedestrian is constant. To get real pace of each pedestrian, the time for 10 steps is measured 25 times using video tape for each pedestrian. Comparison of the real pace and the estimated pace are summarized in **Table 2**. The estimated paces are almost valid with consideration that the pace of pedestrian is not strictly constant.

Table 2: Comparison of (a): real pace and (b): estimated pace by the proposed method (Hz).

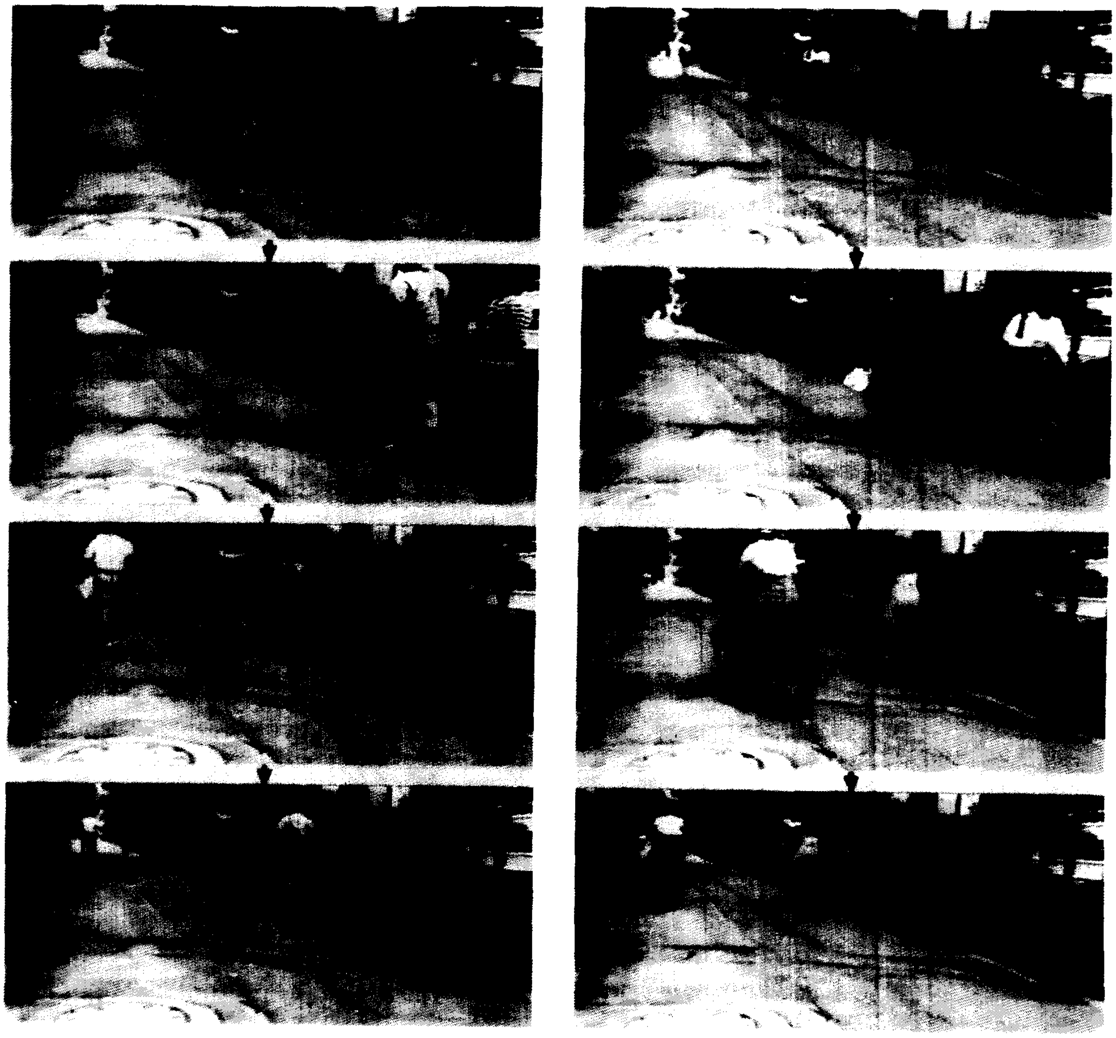
Obj. num.	(1)	(2)	(3)	(4)	(5)	(6)
(a)	1.90	1.92	1.99	1.87	1.86	1.84
(b)	1.92	1.91	1.94	1.86	1.85	1.89

The experimental results of discrimination of 3 hours video tape are summarized in **Table 3**. It shows the result of matching for 231 pedestrians and 22 non-pedestrians (18 bicycles and 4 dogs). Most of the bicycles are not tracked as their width are not in the pedestrian width range. In addition, when the tracking process does not succeed in more than N cycles, the rhythm detection and the model matching is not performed. Therefore, the result of the table does not include such cases.

Table 3: Result of discrimination.

Kind of moving object	Sample number	Result of discrimination	
		Pedestrian	Non-pedestrian
Pedestrian	231	212 (91.8%)	19 (8.2%)
Bicycle	18	0	18 (100%)
Dog	4	0	4 (100%)

The result of experiments indicates that the 8.2% of pedestrians are not judged as Pedestrian. The main reasons of these failures are jitter, occlusion and carrying big



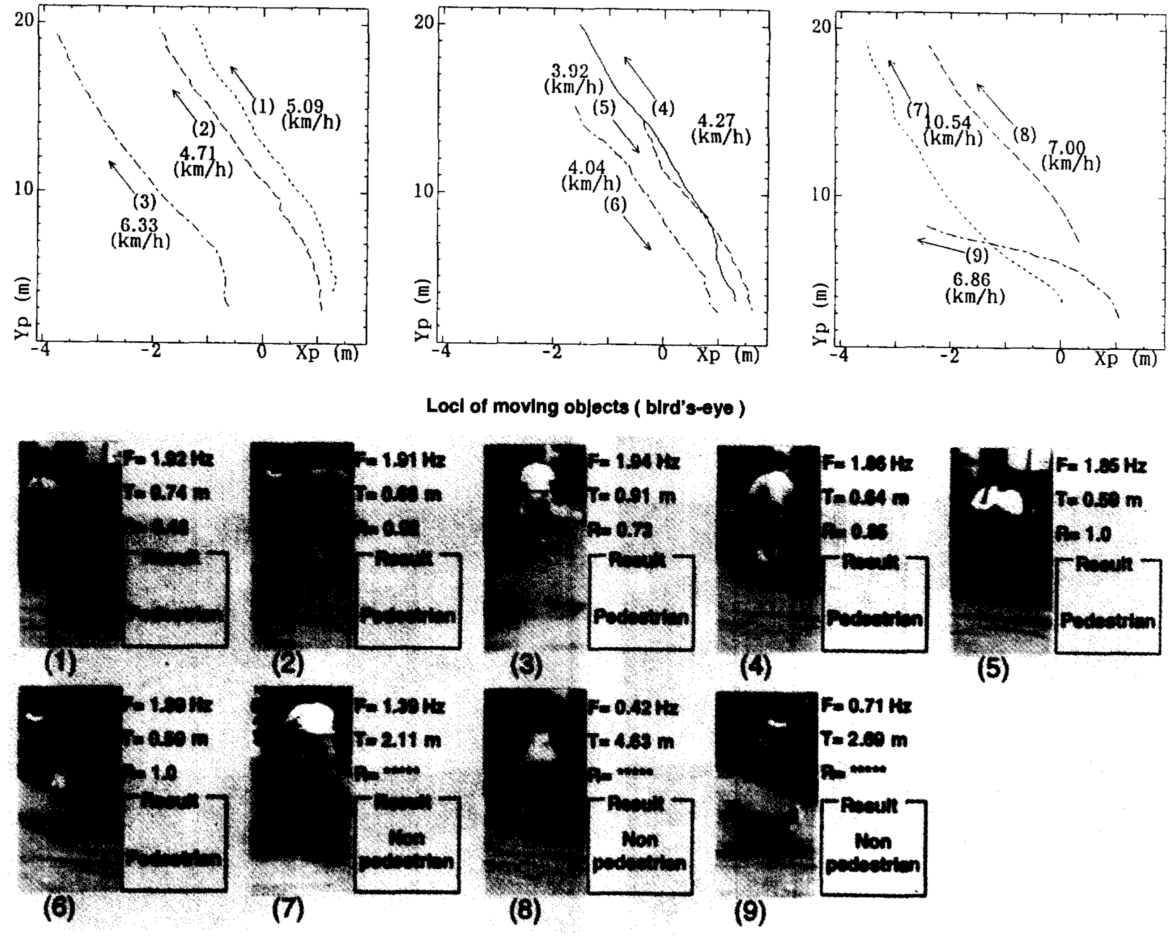
(1) (2)

Figure 5: Image sequences for the experiments.

baggage. In particular, the effect of jitters on the subtracted image is remarkable.

6 Conclusion

We proposed a method of pedestrian discrimination from the other moving objects in the asphalt paved road environment. This method consists of three processes; moving object detection, tracking and discrimination based on rhythm. The moving object detection is performed by subtraction of successive frames. In the tracking, the estimation of moving object position and velocity is implemented by a kinematic model, a measurement model and a tracking filter. The periodical changing of image intensity is detected on subtracted images. Rhythm, that is, the temporal-frequency and the spatial-frequency are used as the features of the periodical motion of walking to discriminate the pedestrians from the other moving objects. This method has some advantages; (1) the rhythm is not affected so much by the kind of clothes, the distance between the object and the observer and weather; (2) the algorithm is simple enough to perform in real time. We implemented this method and showed the ability of pedestrian discrimination from the other moving objects.



Loci of moving objects (bird's-eye)

Figure 6: Experimental results of each object.

The attributes, locus, velocity, pace and stride, are obtained as the features of pedestrian. Experiments were done on the asphalt paved road scenes included of more than 200 pedestrians and the other objects such as bicycles and dogs. 91.8% of pedestrians were successfully discriminated and 100% of the other objects were successfully detected as Non-pedestrian. In addition, the method can be used to get the rough information of the pedestrian before applying a relatively high level identification method.

Acknowledgements

This work has been partially supported by Grant-in-Aid for Scientific Research from Ministry of Education No. 03650343, Okawa Institute of Information and Telecommunication No.91-32 and OMRON Co.Ltd.

References

[1] J. O'Rourke, N. I. Badler, "Model-Based Image Analysis of Human Motion Using Constraint Propagation," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-2, no. 6, pp.522-536, 1980.

[2] M. A. Turk, A. P. Pentland, "Face Recognition Using Eigenfaces," Proc. IEEE Computer Soc. Computer Vision and Pattern Recognition, Lahaina, Maui, Hawaii, June 1991.

[3] G. Johansson, "Spatio-Temporal Differentiation and Integration in Visual Motion Perception," Psych. Res., vol. 38, pp.379-383, 1976.

[4] L. T. Kozlowski, J. E. Cutting, "Recognizing the Sex of a Walker from a Dynamic Point-Light Display,"

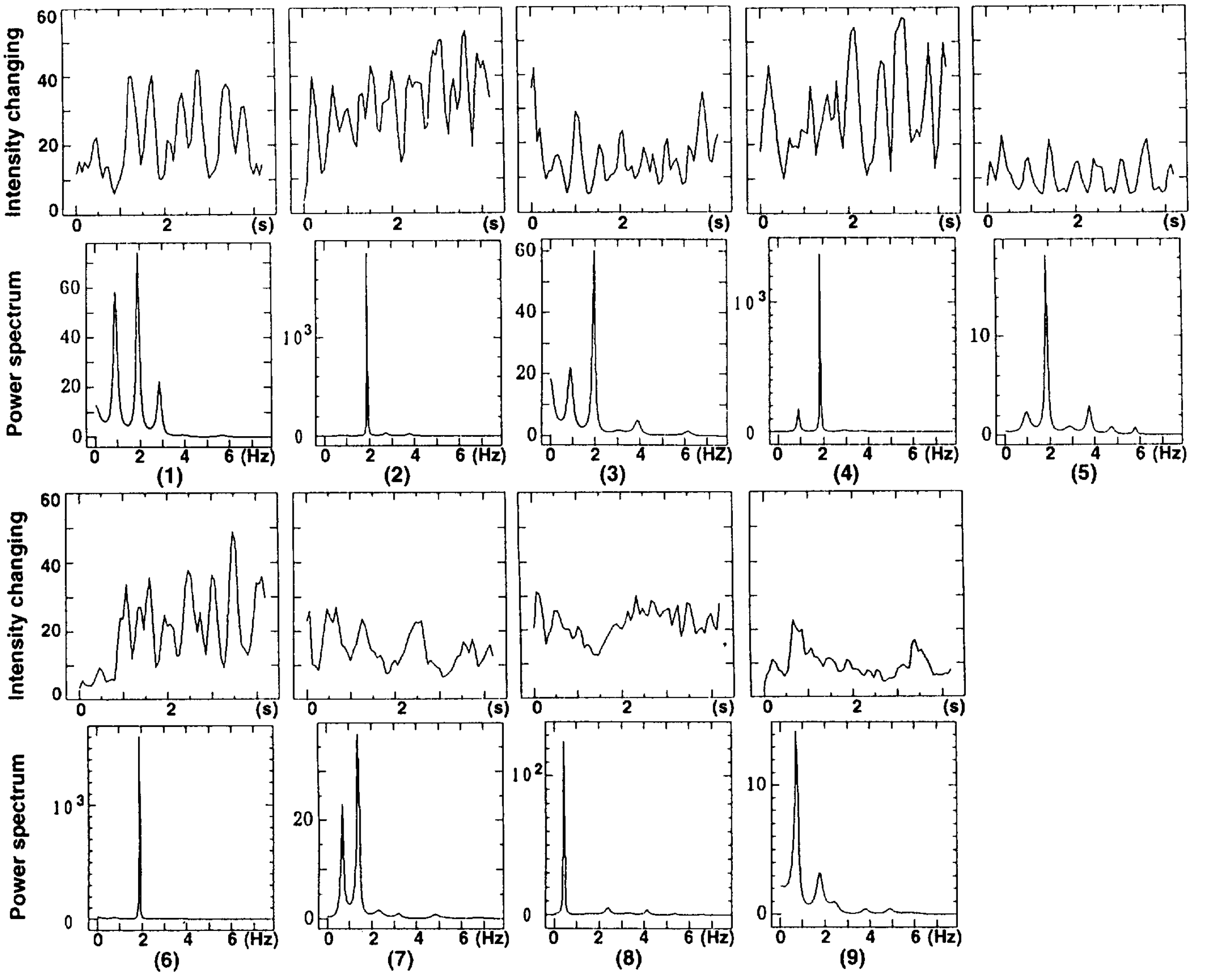


Figure 7: Intensity changing and power spectrum of each object.

Perception & Psychophysics, vol. 21, no. 6, pp.575-580, 1977.

[5] N. Tinbergen, "The Study of Instinct," Oxford University Press, 1951 (in Japanese, Translation: T. Nagano, chapter 2, Sankyo-Shuppan, 1975).

[6] H. Mori, N. M. Charkari, "Shadow and Rhythm as Sign Patterns of Obstacle Detection," Proc. IEEE Int. Symposium Ind. Electronics, pp.271-277, Budapest, Hungary, June 1993.

[7] M. Yachida, M. Asada, S. Tsuji, "Automatic Analysis of Moving Images," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-3, no. 1, pp.12-20, 1981.

[8] A. Shio, J. Sklansky, "Segmentation of People in Motion," Proc. IEEE Workshop on Visual Motion, pp.325-333, 1991.

[9] A. L.' Gilbert, M. G. Giles, G. M. Flachs, R. B. Rogers, Y. H. U, "A Real-Time Video Tracking System," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-2, no. 1, 1980.

[10] G. R. Legters, T. Y. Young, "A Mathematical Model for Computer Image Tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-4, no. 6, pp.583-594, 1982.