# Counting Pedestrians Passing through a Line in Video Sequences based on Optical Flow Extraction 

MIKI MIZUSHIMA, YOSHIAKI TANIGUCHI, GO HASEGAWA, HIROTAKA NAKANO, and MORITO MATSUOKA<br>Graduate School of Information Science and Technology Osaka University<br>1-5 Yamadaoka, Suita-shi, Osaka<br>JAPAN<br>\{m-mizusm, y-tanigu, hasegawa, nakano, matsuoka\}@ist.osaka-u.ac.jp


#### Abstract

In recent researches on video processing-based pedestrian counting, a pedestrian density estimation method based on the number of optical flows has developed and attracted attentions. In this paper, we propose a simple pedestrian counting method by applying the number of optical flows for estimating the number of pedestrians passing through a specific line, called counting line, in crowded situations. Through experimental evaluations using actual video sequences, we show that the correlation coefficient between the number of optical flows and the number of pedestrians passing through the counting line is high. In addition, we show that the mean relative error of our proposed method is less than 0.068


Key-Words: pedestrian counting, video sequence, optical flow, counting line, good feature, image processing

## 1 Introduction

Pedestrian counting is needed in many situations such as safety control in airports and stations, traffic control in streets, and surveillance at event sites. For example, the information on the number of passengers in a station or an airport is useful to give an appropriate reaction in the case of emergency or disruption of services. In addition, information on the number of people in a office building or a public facility can be used for safety management and energy saving control of buildings [1]. These information can be obtained by monitoring the number of pedestrians passing through specific areas such as doors, passageways, and so on. Pedestrian counting is usually conducted manually or automatically using sensors or video cameras. Manual counting requires labor costs and is difficult to apply in crowded scenes due to human errors. Therefore, automatic counting has attracted a lot of attention.

Among automatic pedestrian counting methods, video processing-based counting methods have been widely studied. Although the performance of video proceeding-based counting is affected by surrounding environments such as the changes of lights and the weather, it has some advantages. For example, camera placement is flexible because video sequence can be retrieved remotely. In addition, the introduction cost can be reduced by using existing surveillance cameras for some situations. Therefore, there have been many researches for counting pedestrians in video se-
quences.
Pedestrian counting methods can be divided to pedestrian detection-based methods (also known as direct methods) and feature-based methods (also known as indirect methods, map-based methods, measurement-based methods) [2, 3]. Pedestrian detection-based methods [4-9] are the traditional methods to count the number of pedestrians. In these methods, the regions of pedestrians are detected from a frame, and the number of regions is counted. A lot of methods for detecting regions of pedestrians have been proposed previously as summarized in [6], and they can be used for the counting purpose. However, their counting accuracy significantly decreases in crowded situation where the density of pedestrians is high and occlusion occurs frequently.

Recently, pedestrian counting in crowded scenes has attracted attention and discussed in some conferences or workshops such as IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS). In the crowded situations, instead of pedestrian detection-based methods, featurebased methods [1-3, 10-17] have become popular for pedestrian counting. In feature-based methods, various features, such as the amount of moving pixels [2, 11], the number of optical flows [3, 12], are extracted from video sequences to estimate the number of pedestrians. The number of pedestrians is estimated based on the number or the amount of fea-
tures and pre-learned correlation between them and the number of pedestrians. Feature-based methods have benefits in terms of simplicity, processing speed, and so on.

In the viewpoint of purpose, video processingbased pedestrian counting methods can be classified into two kinds of purposes. One is for counting pedestrians in a specific area or whole area in a frame [1, 3, $4,8,10,12-14,18,19]$. These methods are also called as pedestrian density estimation methods. In [12], the authors proposed a simple feature-based pedestrian density estimation method. In the method, feature points are first extracted from video sequences. Then, optical flows whose starting points are the feature points are obtained. Pedestrian density is estimated based on the number of optical flows whose length are not zero. In the PETS 2009 workshop, the method shows good performance among methods presented at the workshop under evaluations using same benchmark video sequences. Our research group also proposed a feature-based pedestrian density estimation method [13] inspired by the method in [12]. In the method, optical flows are clustered and pedestrian density is estimated based on the number of clusters.

On the other hand, the other kind of methods is for counting pedestrians passing through a specific line or a specific area $[2,9,15-17]$. In this paper, we call the line counting line. For example, the counting line is set at escalators, an entrance of buildings, and so on. Researches on counting pedestrians passing through a counting line or a counting area have been less studied compared to researches on pedestrian density estimation. Some existing methods are the feature-based methods [2, 15, 17]. In these methods, the combination of blob size and the number of optical flows in a blob [2], feature point trajectories [15], or foreground pixels [17] are used for counting the number of pedestrians. However, these methods are complex, used a variety of features, or targeted to count the number of pedestrians passing through a counting area. Simple methods for counting pedestrians passing through a counting line have not been discussed sufficiently.

In this paper, we apply the number of optical flows to estimate the number of pedestrians passing through a counting line. We consider a crowded situation where the density of pedestrians in each frames is high and occlusion occurs frequently. In our proposed method, to count pedestrians, feature points are firstly detected from video sequences. Next, optical flows whose starting points are the feature points are extracted as in [12, 13]. Then, the number of pedestrians passing through the counting line is estimated based on the number of optical flows. In this paper, the performance of our proposed method is evaluated using actual video sequences.


Figure 1: Pedestrian counting system

The rest of this paper is organized as follows. Section 2 explains the overview of our counting system intended in this paper. Section 3 explains our proposed counting method based on optical flow extraction. Section 4 discusses experimental results based on actual video sequences. Finally, Section 5 concludes this paper with outlook of future work.

## 2 System Overview

The overview of our counting system intended in this paper is shown in Fig. 1. A user first places a stationary camera to obtain video sequences of the monitoring area where pedestrians should be count. Video sequences are recorded and displayed at a monitoring server. Then, the user sets a counting line in the video sequences. The counting line is set at a location where the user wants to measure the number of pedestrians, for example, a doorway or a passageway of the buildings. In this paper, we assume only one counting line for simplicity.

The monitoring server estimates the number of pedestrians passing through the counting line until a certain frame, and outputs the information to the user. In the next section, we propose a method for counting the number of pedestrians passing through the counting line.

## 3 Proposed Method

In this section, we propose a method for estimating the number of pedestrians passing though the counting line. In our proposed method, feature points are
firstly detected from video sequences as start points of optical flows. Then, optical flows which start from the feature points are extracted. Finally, the number of pedestrians is estimated based on the number of optical flows using the pre-learned correlation between the number of optical flows and the actual number of pedestrians as in [12].

In the following, the $i$-th frame, the $j$-th feature point in the $i$-th frame, and the optical flow starting from the $j$-th feature point in the $i$-th frame are denoted as $f_{i}, p_{i, j}$, and $o_{i, j}$, respectively. The coordinate of start point and the end point of the counting line are denoted as $\left(x_{g a}, y_{g a}\right),\left(x_{g b}, y_{g b}\right)$, respectively.

### 3.1 Detecting Feature Points

In our proposed method, feature points are detected from moving parts in the frame as in [13]. We first identify the moving parts in the frame by using the background difference method [20]. In the method, to mitigate the adverse effects of changes in the illumination conditions, the exponential weighted moving average calculation is used to obtain pixel values of background frame as follows:

$$
\begin{equation*}
\bar{w}_{i, x, y}=(1-\mu) \bar{w}_{i-1, x, y}+\mu w_{i, x, y} \tag{1}
\end{equation*}
$$

Here, $w_{i, x, y}$ denotes the pixel value at coordinate ( $x$, $y$ ) of frame $f_{i}, \bar{w}_{i, x, y}$ denotes the pixel value of the background frame at coordinate $(x, y)$, and $\mu$ is a weight parameter. The moving parts in the frame are identified by evaluating the difference between the pixel value of the current frame $w_{i, x, y}$ and that of the background frame $\bar{w}_{i, x, y}$. If the difference is larger than threshold value $w_{d}$, the pixel at coordinate $(x, y)$ is considered as in the moving parts of frame $f_{i}$.

Then, we extract feature points from the moving parts by using the Shi-Tomasi corner detection method [21]. In the method [21], eigenvalues for each pixel value in the frame are calculated. Then, feature points are selected from the point where the eigenvalue is above a threshold value. We denote the set of feature points in frame $f_{i}$ as $\mathcal{P}_{i}=$ $\left\{p_{i, 1}, p_{i, 2}, \cdots, p_{i, r_{i}}\right\}$ where $r_{i}$ is the number of feature points in frame $f_{i}$.

### 3.2 Extracting Optical Flows

We next extract optical flows based on the detected feature points in the previous section. The set of optical flows, $\mathcal{O}_{i}=\left\{o_{i, 1}, o_{i, 2}, \cdots\right\}$, is extracted from the set of feature points $p_{i, j} \in \mathcal{P}_{i}$, in frame $f_{i}$ by estimating each corresponding coordinate in frame $f_{i+1}$. In this paper, we use the Lucas-Kanade method [22] to extract optical flows from the feature points. In the
method [22], optical flows can be estimated under the following assumptions.

- The brightness of each moving part remains unchanged in frames $f_{i}$ and $f_{i+1}$.
- The distance traveled by a moving part between frames $f_{i}$ and $f_{i+1}$ is within a certain range.
- Neighboring points such as $\left(x_{i}, y_{i}\right)$ and $\left(x_{i}+1\right.$, $y_{i}+1$ ) of frame $f_{i}$, which belong to the same moving part, retain the same relative positions in frame $f_{i+1}$.

We call the line segment from the starting point of an optical flow to the end point of the optical flow as optical flow segment.

Incorrect optical flows are sometimes extracted due to noise in video sequences. In this paper, we delete incorrect optical flows from the set of optical flows $\mathcal{O}_{i}$ based on the length of optical flow. The length of the extracted optical flow $o_{i, j} \in \mathcal{O}_{i}$ is calculated by Eqn. (2).

$$
\begin{equation*}
l_{i, j}=\sqrt{\left(x_{i, j}^{e}-x_{i, j}^{s}\right)^{2}+\left(y_{i, j}^{e}-y_{i, j}^{s}\right)^{2}} \tag{2}
\end{equation*}
$$

Here, $\left(x_{i, j}^{s}, y_{i, j}^{s}\right)$ and $\left(x_{i, j}^{e}, y_{i, j}^{e}\right)$ denotes the coordinate of the start point and that of end point of optical flow $o_{i, j}$. If the length $l_{i, j}$ does not satisfy Eqn. (3), we delete optical flow $o_{i, j}$ from the set of optical flows $\mathcal{O}_{i}$.

$$
\begin{equation*}
l_{\min }<l_{i, j}<l_{\max } \tag{3}
\end{equation*}
$$

$l_{\min }$ and $l_{\max }$ are the minimum and the maximum thresholds to determine the acceptable length of optical flows. These values should be appropriately configured preliminarily.

Since the number of pedestrians passing through the counting line should be estimated, we choose a part of optical flows for utilizing to estimate the number of pedestrians. If both Eqns. (4) and (5) are satisfied for optical flow $o_{i, j}$, the optical flow segment intersects the counting line.

$$
\begin{align*}
& \left(\left(x_{g a}-x_{g b}\right)\left(y_{i, j}^{e}-y_{g a}\right)+\left(y_{g a}-y_{g b}\right)\left(x_{g a}-x_{i, j}^{e}\right)\right) \\
& \left(\left(x_{g a}-x_{g b}\right)\left(y_{i, j}^{s}-y_{g a}\right)+\left(y_{g a}-y_{g b}\right)\left(x_{g a}-x_{i, j}^{s}\right)\right) \\
& \leq 0 .  \tag{4}\\
& \left(\left(x_{i, j}^{e}-x_{i, j}^{s}\right)\left(y_{g a}-y_{i, j}^{e}\right)+\left(y_{i, j}^{e}-y_{i, j}^{s}\right)\left(x_{i, j}^{e}-x_{g a}\right)\right) \\
& \left(\left(x_{i, j}^{e}-x_{i, j}^{s}\right)\left(y_{g b}-y_{i, j}^{e}\right)+\left(y_{i, j}^{e}-y_{i, j}^{s}\right)\left(x_{i, j}^{e}-x_{g b}\right)\right) \\
& \leq 0 . \tag{5}
\end{align*}
$$

Otherwise, the optical flow segment does not intersect the counting line and the optical flow $o_{i, j}$ is deleted from the set of optical flows $\mathcal{O}_{i}$.

### 3.3 Estimating the Number of Pedestrians

We use optical flows obtained from the previous section for counting pedestrians passing through the counting line. The total number of pedestrians passing through the counting line from the start frame to frame $f_{i}$ is estimated as follows:

$$
\begin{equation*}
\hat{n}_{i}=\left\lfloor\rho \sum_{k=1}^{i} c_{k}\right\rfloor \tag{6}
\end{equation*}
$$

where $c_{k}=\left|\mathcal{O}_{k}\right|$ denotes the number of optical flows in frame $f_{k}$, and $\rho$ is a parameter that indicates the average number of pedestrians per one optical flow. Parameter $\rho$ is preliminary obtained from a training video data set as follows.

$$
\begin{equation*}
\rho=\frac{n_{d_{l r n}}}{\sum_{i=1}^{d_{l r n}-1} c_{i}} \tag{7}
\end{equation*}
$$

where $n_{d_{l r n}}$ is the number of pedestrians in the training video and $d_{l r n}$ is the number of frames in the training video.

## 4 Experiments and Evaluation

### 4.1 Experimental Environment

To confirm the fundamental performance of our proposed method, we conduct experiments and evaluations as follows. We use three video sequences as summarized in Table 1. Video 1 and video 2 were recorded at the Cybermedia Center in Osaka University, Japan. In video 1 , ten pedestrians move bidirectionally in a pathway. On the other hand, around twenty pedestrians move bi-directionally in a pathway in video 2 . In video 2 , the density of the pedestrians and occurrence frequency are higher than that in video 1. Video 3 was recorded at a crowded crosswalk in Tokyo, Japan. The density of the pedestrians and occurrence frequency of video 3 are the highest among all video sequences. We set a counting line for each video sequence as shown in Fig. 2. In all video sequences, pedestrians move bi-directionally against the counting line. The actual number of pedestrians passing through the counting line is measured manually when the center of pedestrian's body passed the counting line. Our proposed method is implemented on an off-the-shelf PC (Intel Core i7-2600 CPU 3.40 [GHz], 8.00 [GB] memory, Microsoft Windows 7 Professional). Table 2 shows the values of parameter used in this paper.

As evaluation metrics of our proposed method, we use the average of mean absolute error (MAE) and


Figure 2: Evaluation video sequences (blue line is the counting line)

Table 1: Summary of video sequences

|  | Video 1 | Video 2 | Video 3 |
| :--- | :---: | :---: | :---: |
| Place | Osaka Univ | Osaka Univ | Crosswalk |
| The total <br> number of <br> pedestrians | 10 | 19 | 34 |
| Frame rate | 29.97 | 29.97 | 29.97 |
| The total <br> number of <br> frames | 640 | 640 | 640 |
| Resolution | $720 \times 480$ | $720 \times 480$ | $720 \times 480$ |
| The average <br> number of <br> optical flows <br> per one <br> pedestrian | 9.80 | 8.74 | 10.23 |

mean relative error (MRE) as follows:

$$
\begin{aligned}
& M A E=\frac{1}{d_{\text {end }}-d_{\text {start }}} \sum_{i=d_{\text {start }}}^{d_{\text {end }}}\left|\hat{n}_{i}-n_{i}\right| \\
& M R E=\frac{1}{d_{\text {end }}-d_{\text {start }}} \sum_{i=d_{\text {start }}}^{d_{\text {end }}} \frac{\left|\hat{n}_{i}-n_{i}\right|}{n_{i}} .
\end{aligned}
$$

Here, $d_{\text {start }}$ and $d_{\text {end }}$ are the first and the last frame of video sequences for evaluation, respectively. $\hat{n}_{i}$ is the total number of estimated pedestrians until frame $f_{i}$ and $n_{i}$ is the total number of actual pedestrians until frame $f_{i}$. Smaller MAE and MRE indicate higher estimation accuracy.

### 4.2 Investigation of the Number of Optical Flows

In this section, we first investigate the number of optical flows obtained from each pedestrians to know fundamental characteristics to use optical flows for pedestrian counting. We obtained the number of optical flows manually from the video sequences.

Figure 3 shows the histogram of the number of optical flows obtained from each pedestrian passing

Table 2: Parameter settings

| $\mu$ | 0.05 |
| :---: | :---: |
| $w_{d}$ | 128 |
| $l_{\min }$ | 0.2 [pixel] |
| $l_{\max }$ | 25.0 [pixel] |



Figure 3: Histogram of the number of optical flows obtained from a pedestrian

Table 3: Correlation coefficient between the number of optical flows and the number of pedestrians, and estimation errors of our proposed method

|  | Video 1 | Video 2 | Video 3 |
| :---: | :---: | :---: | :---: |
| Correl | 0.891 | 0.809 | - |
| MAE | 0.680 | 0.652 | 1.153 |
| MRE | 0.068 | 0.034 | 0.034 |

through the counting line. As shown in Fig. 3, the shape of histogram is slightly convex with a broad range of the number of optical flows. Since the number of feature points form a pedestrian is affected by some factors such as clothes, bags, and occlusion, the number of optical flows obtained from a pedestrian varies. In video 2 and video 3 , since the density of the pedestrians is high and occlusion frequently occurs, the number of optical flows obtained from a pedestrian highly varies compared to that in video 1 . To suppress the variation, one possible way is to sum up similar optical flows to one optical flow by clustering. The detailed discussion is our future work.

We next investigate the relationship between the number of optical flows and the total number of pedestrians in a group passing through the counting line in a part of video sequence. To investigate the relationships, we divide each video sequence to sub-video sequences such that the video sequence is cut when there is no moving parts on the counting line. For each subvideo sequence, we obtain the number of optical flows and the number of pedestrians. Then, we calculate the correlation coefficient for each video sequence.

Table 3 shows the correlation coefficient between the number of optical flows and the number of pedestrians. We do not show the results of video 3, since the video sequence can not be divided to sub-video sequences. As shown in Table 3, the correlation coefficients of video 1 and 2 are above 0.8 and the number of optical flows and the number of pedestrians are highly correlated, although the number of optical flows obtained from each pedestrian highly varies as described in the previous paragraph. Since occlusion occurs infrequently in video 1 , the correlation coefficient of video 1 is higher than that of video 2. From these results, we conclude that it is reasonable to use the number of optical flows to estimate the number of pedestrians.

### 4.3 Evaluation of Counting Accuracy

We next evaluate the counting accuracy of our proposed method. To evaluate the fundamental perfor-


Figure 4: The total number of pedestrians passing through the counting line
mance of our proposed method, we used each whole video sequence as a training video.

Figure 4 shows comparisons between the actual number of pedestrians and estimated one. In Fig. 4 $x$-axis corresponds to the number of frames and $y$ axis corresponds to the number of pedestrians. Table 3 shows MAE and MRE of our proposed method. The comparison of MAE and MRE in video 1 and video 2 shows that MAE and MRE in video 1 are higher although the correlation coefficient in video 1 is higher as described in the previous section. This is because counting the first pedestrian passing through the counting line in video 1 is failed in our proposed method at around frame 30, and no additional pedestrian passes the counting line until frame 270 as shown in Fig. 4. Since the number of optical flows from the first pedestrian is fewer than the average number of optical flows per pedestrian, no pedestrian is counted until frame 270. To improve the counting accuracy in the situation, one possible way is to use round-off in Eqn. (6) instead of round-down. Detailed discussion and evaluation on Eqn. (6) is one of our future work.

In video 3, MAE of our proposed method is the highest among the evaluated video sequences since the number of pedestrians is highest among the video sequences. However, MRE of our proposed method is less than or equal to that in other video sequences, although the scene in video 3 is crowded. In this evaluation, our proposed method is capable of counting pedestrians in crowded situations.

## 5 Conclusion and Future work

In this paper, we applied the number of optical flows to estimate the number of pedestrians in crowded situations. We proposed a simple method for counting pedestrians passing through a counting line. Through experimental evaluations using actual video sequences, we showed that the correlation coefficient between the number of optical flows and the number of pedestrians is high. In addition, we showed that the mean relative error of our proposed method is less than 0.068 .

As future works, we should evaluate our proposed method in comparison with other pedestrian counting methods. We also plan to improve the accuracy of our proposed method by introducing an optical flow clustering method as in [13]

## References:

[1] H. Fradi and J. Dugelay. Low level crowd analysis using frame-wise normalized feature for people counting. In Proceedings of IEEE WIFS 2012, pp. 246-251, December 2012.
[2] Y. Benabbas, N. Ihaddadene, T. Yahiaoui, T. Urruty, and C. Djeraba. Spatio-temporal optical flow analysis for people counting. In Proceedings of IEEE AVSS 2010, pp. 212-126, August 2010.
[3] D. Conte, P. Foggia, G. Percannella, F. Tufono, and M. Vento. A method for counting people in crowded scenes. In Proceedings of IEEE AVSS 2010, pp. 225-232, August 2010.
[4] G. J. Brostow and R. Cipolla. Unsupervised bayesian detection of independent motion in crowds. In Proceedings of IEEE CVPR 2006, pp. 594-601, June 2006.
[5] Y. C. Zeng S. W. Sun D. Y. Chen, C. W. Su and H. Y. M. Liao. An online people counting system for electronic advertising machines. In Proceedings of ICME 2009, pp. 1262-1265, July 2009.
[6] M. Enzweilar and D. M. Gavrila. Monocular pedestrian detection: Survey and experiments. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 12, pp. 21792195, December 2009.
[7] W. Ritter D. Paulus R. Arndt, R. Schweiger and O. Lohlein. Detection and tracking of multiple pedestrians in automotive applications. In Proceedings of Inteligent Vehicles Symposium 2007, pp. 13-18, June 2007.
[8] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In Proceedings of ECCV 2006, pp. 428-441, May 2006.
[9] C. Chen T. Chen and D. Wang. A cost-effective people-counter for passing through a gate based on image processing. In Proceedings of ICIC 2009, pp. 785-800, March 2009.
[10] G. Antonini and J.P. Thiran. Counting pedestrians in video sequences using trajectory clustering. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 16, No. 8, pp. 10081020, August 2006.
[11] Z. S. J. Liang A. B. Chan and N. Vasconcelos. Privacy preserving crowd monitoring: Counting
people without people model or tracking. In Proceedings of IEEE CVPR 2008, pp. 1-7, June 2008.
[12] A. Albiol, M. J. Silla, A. Albiol, and J. M. Mossi. Video analysis using corner motion statistics. In Proceedings of IEEE PETS 2009, pp. 31-37, June 2009.
[13] S. Fujisawa, G. Hasegawa, Y. Taniguchi, and H. Nakano. Pedestrian counting in video sequences based on optical flow clustering. International Journal of Image Processing, Vol. 7, No. 1, pp. 1-16, February 2013.
[14] H. Fradi and J. Dugelay. People counting system in crowded scenes based on feature regression. In Proceedings of EUSIPCO 2012, pp. 136-140, August 2012.
[15] D. Lamovsky and R. Sadykhov. Method of pedestrians traffic assessment based on analysis of video data in surveillance systems. In Proceedings of MIPRO 2010, pp. 704-706, May 2010.
[16] T. Yang, Y. Zhang, D Shao, and Y. Li. Clustering method for counting passengers getting in a bus with single camera. Optical Engineering, Vol. 49, No. 3, pp. 1-10, March 2010.
[17] K. Ng S. Srivastava and E. Delp. Crowd flow estimation using multiple visual features for scenes with changing crowd densities. In Proceedings of IEEE AVSS 2011, pp. 246-251, August 2011.
[18] E. Bas, A. M. Tekalp, and F. S. Salman. Automatic vehicle counting from video for traffic flow analysis. In Proceedings of IEEE IVS 2007, pp. 392-398, June 2007.
[19] Y. Hou and G. Pang. Automated people counting at a mass site. In Proceedings of IEEE ICAL 2008, pp. 464-469, September 2008.
[20] A. Elegammal, D. Harwood, and L. Davis. Nonparametric model for background subtraction. In Proceedings of IEEE CVPR 2000, pp. 751-767, June 2000.
[21] J. Shi and C. Tomasi. Good features to track. In Proceedings of IEEE CVPR 1994, pp. 593-600, June 1994.
[22] A. Elegammal, D. Harwood, and L. Davis. An iterative image registration technique with an application to stereo vision. In Proceedings of the

1981 DARPA Image Understanding Workshop, pp. 121-130, April 1981.

