Pedestrian Counting in Video Sequences based on Optical Flow Clustering

Shizuka Fujisawa s-fujisw@ist.osaka-u.ac.jp
Graduate School of Information Science and Technology
Osaka University
Suita, Osaka, 565-0871, Japan

Go Hasegawa hasegawa@cmc.osaka-u.ac.jp
Cybermedia Center
Osaka University
Toyonaka, Osaka, 560-0043, Japan

Yoshiaki Taniguchi y-tanigu@cmc.osaka-u.ac.jp
Cybermedia Center
Osaka University
Toyonaka, Osaka, 560-0043, Japan

Hirotaka Nakano nakano@cmc.osaka-u.ac.jp
Cybermedia Center
Osaka University
Toyonaka, Osaka, 560-0043, Japan

Abstract

The demand for automatic counting of pedestrians at event sites, buildings, or streets has been increased. Existing systems for counting pedestrians in video sequences have a problem that counting accuracy degrades when many pedestrians coexist and occlusion occurs frequently. In this paper, we introduce a method of clustering optical flows extracted from pedestrians in video frames to improve the counting accuracy. The proposed method counts the number of pedestrians by using pre-learned statistics, based on the strong correlation between the number of optical flow clusters and the actual number of pedestrians. We evaluate the accuracy of the proposed method using several video sequences, focusing in particular on the effect of parameters for optical flow clustering. We find that the proposed method improves the counting accuracy by up to 25% as compared with a non-clustering method. We also report that using a clustering threshold of angles less than 1 degree is effective for enhancing counting accuracy. Furthermore, we compare the performance of two algorithms that use feature points and lattice points when optical flows are detected. We confirm that the counting accuracy using feature points is higher than that using lattice points especially when the number of occluded pedestrians increases.

Keywords: Pedestrian Counting, Video Sequence, Optical Flow, Clustering.

1. INTRODUCTION

Pedestrian counting at event sites, buildings, or streets is an essential function for pedestrian traffic control, marketing surveys, surveillance in public places, and so on. For example, counting customers who enter and exit from each shop in shopping mall can be used for measuring business performance of the shops. Counting pedestrians at streets is effective to decide when to do road construction and maintenance. Traditionally, pedestrian counting is conducted manually, that has some disadvantages. For example, personnel expenses are high, simple work for long time is a cause of stress, it may introduce human errors, and so on. Therefore, automatic pedestrian counting has received much attention [1], [2].

Numerous approaches to automatically counting moving pedestrians have been reported in the literature. The most widely deployed methods utilize laser sensors [3], [4], [5], [6] and infrared sensors [7], [8], [9], [10], [11], [12], [13]. For example, [6] proposes the counting pedestrians in real time using a network of single row laser range scanners. Such sensor-based methods have advantages in terms of low cost and robustness against changes in the illumination conditions due to fluctuations of lighting and weather. However, these methods sometimes fail to count pedestrians correctly when the heights of pedestrians walking together are similar or when the heights do not fall within the presumed range, because such methods depend on the difference in propagation delays of reflected laser pulses or infrared light. Although the methods using multiple infrared sensors [7], [8], [9] can count pedestrians moving various directions, the counting accuracy degrades considerably when the street has much traffic and occlusion occurs frequently [14]. An occlusion is caused by pedestrians interacting with each other when many pedestrians are present [15].

On the other hand, pedestrian counting methods using video processing technologies have some advantages. For instance, the camera position is flexible such that it is vertically and obliquely above the pedestrians. Additionally, the video sequences can be retrieved at the distanced position from the target area where pedestrians are to be counted. Furthermore, information such as the height and gender of pedestrians can be obtained, in addition to the number of pedestrians. However, the methods using video processing technologies are affected by changes in the illumination conditions [16]. Furthermore, the counting accuracy of these methods degrades due to occlusion effect [17].

The counting pedestrians by video processing technologies is being actively researched, and a contest for counting pedestrians by using video processing technologies is held at the annual IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS) [18]. Texture-based methods [19], [20], [21], [22] are robust to the occlusion of moving objects whose texture pattern is uniform. However, people generally have non-uniformity in their appearance. Model-based methods [15], [23], [24], [25], [26] use a shape model to extract moving objects, and have the high counting accuracy in the presence of occlusions. However, the performance of these methods becomes considerably worse when the moving object is a non-rigid object such as a person. In [27], [28], a camera is set directly above pedestrians to avoid occlusion. However, the camera position is restricted to locations such as the entrance of a building. The methods in [29], [30], [31] use multiple cameras which give the three-dimensional positions of pedestrians to identify their locations more accurately. However, computing three-dimensional positions from point correspondences over multiple images leads to high computational complexity.

In this paper, we propose a method for counting the number of pedestrians in the video sequences retrieved by a single stationary camera. The proposed method is based on clustering optical flows of moving objects in video frames. Clustering optical flows is important for distinguishing multiple pedestrians who are walking together. Since optical flows detected from a single pedestrian have similar lengths, angles, and since their starting points exist within a certain range, we utilize such parameters in clustering. The proposed method estimates the number of pedestrians based on the strong correlation between the number of optical flow clusters detected by proposed method and the actual number of pedestrians. Since the degree of correlation depends on camera position and various other factors, we utilize statistical data to estimate the number of pedestrians.

The effectiveness of the proposed method is evaluated through implementation experiments with several video sequences, including the one used at the PETS contest. We compare the performance of the proposed method with a non-clustering method. Additionally, the performance of two kinds of optical flow extraction algorithms is evaluated. We also focus on the relationships between the optimal threshold of optical flow angles in clustering and the upper-limit of angle differentiation resolution of the method.

The rest of this paper is organized as follows. In Section 2, we introduce the overview of the pedestrian counting method in the video sequences. Section 3 explains the detail of optical flow clustering of the method. The pedestrian counting method is evaluated by using several video sequences in Section 4. Finally, Section 5 concludes the paper and describes future work.

2. SYSTEM DESCRIPTION

2.1 Overview

The overview of pedestrian counting method is depicted in Figure 1. The method places a stationary camera at an arbitrary position and records video of the target area (e.g., a street or building entrance) where pedestrians are to be counted. Retrieved video sequences are sent to a server, which computes the estimate of the number of pedestrians.

On the server, the method first uses the background difference method to identify the moving parts in each video frame. Next, optical flows of moving parts are detected by using a well-known method, followed by clustering them based on the lengths, angles, and source locations of detected optical flows. Finally, the number of pedestrians is estimated, based on the pre-learning correlation between the number of optical flow clusters and the actual number of pedestrians. In what follows, each step of the method proposed is explained in turn.

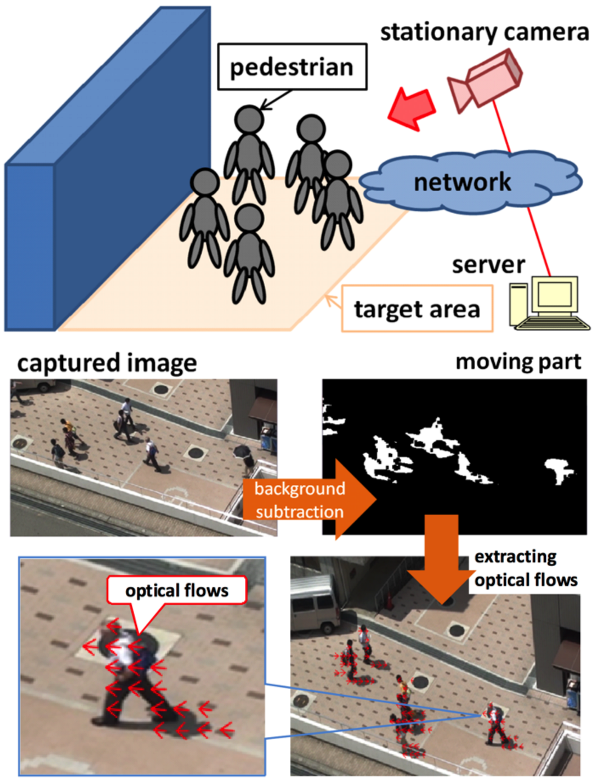


FIGURE 1: Overview of Pedestrian Counting Method.

2.2 Extracting Optical Flows

For the i-th frame ($2 \leq i \leq F$, where F is the total number of frames in the video sequence), denoted by f_i, the method identifies the moving parts in the frame by using the background difference method [32], [33], [34], [35]. To mitigate the adverse effects of changes in the illumination conditions, the method updates the background frame dynamically by using the exponential weighted moving average calculation as follows [36]:

$$\bar{B}_{i,x,y} = (1-\alpha)\bar{B}_{i-1,x,y} + \alpha B_{i,x,y} \tag{1}$$

Here, $B_{i,x,y}$ $(B \in \{R, G, B\})$denotes the RGB pixel value at coordinate $(x, y)$of frame f_i, $\bar{B}_{i,x,y}$ $(\bar{B} \in \{R, G, B\})$ denotes the RGB pixel value at coordinate $(x, y)$of the background frame, and α is a weight parameter. The moving parts are then identified in the frame by evaluating the difference between the RGB pixel value of the current frame and that of the background frame.

$$O_{i,x,y} = \begin{cases} 1 \ (d_{i,x,y} > p_d) \\ 0 \ (d_{i,x,y} \le p_d) \end{cases} \tag{2}$$

$$d_{i,x,y} = \left| B_{i,x,y} - \bar{B}_{i-1,x,y} \right| \tag{3}$$

where $O_{i,x,y}$ indicates whether (x, y) is in moving parts of the frame. When $O_{i,x,y} = 1$, (x, y) is in moving parts of the frame. p_d is the threshold value.

Next, the method places multiple points on the moving parts for extracting optical flows. We consider the following two algorithms and compare them in Subsection 4.4. One uses the *feature points* detected by Harris method [37]. In Harris method, the corners of moving parts, which are defined as the intersection of two edges [38], are used as the points for detecting optical flows, as depicted in Figure 2. The other uses the lattice points, which are placed equally spaced in moving parts, as depicted in Figure 3. Figures 4 and 5 are snapshots where the feature points and the lattice points are detected in a video sequence, respectively.

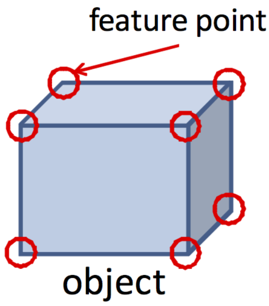


FIGURE 2: Feature Points Detected By Harris Method.

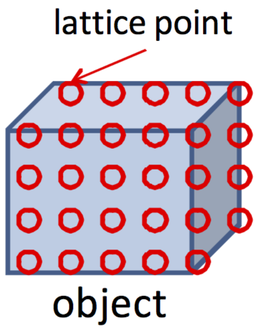


FIGURE 3: Lattice Points.

FIGURE 4: Feature Points Detected In A Video Sequence.

FIGURE 5: Lattice Points in A Video Sequence.

The method then extracts optical flows at multiple points placed using each method in each moving part of the frame. Optical flows are calculated by the Lucas-Kanade method [39]. Here, we denote the coordinate of the j-th point as (x_j, y_j). The Lucas-Kanade method requires the following assumptions.

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

Figure 6 depicts that a point of frame f_{i-1} which corresponds to a point (x_k, y_k) of frame fi as a result of movement is detected. By denoting the coordinate of such a point as (x_j, y_j), the method determines that the optical flow whose starting point is (x_j, y_j) and whose ending point is (x_k, y_k) of frame f_i.

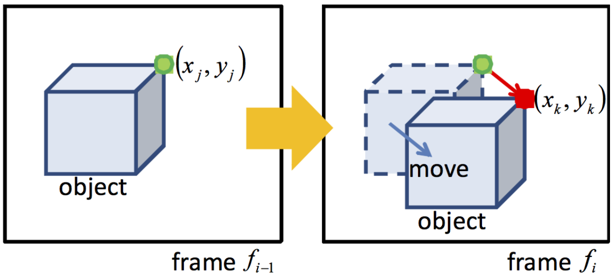


FIGURE 6: Detecting Optical Flows By Lucas-Kanade Method.

To alleviate the effects of noises and changes in the illumination conditions, the method deletes optical flows whose lengths do not fall into a certain range. That is, the method utilizes only the optical flows that satisfy the following condition, where $l_{i,j}$ is the length of an optical flow whose starting point is (x_j, y_j) of frame f_i.

$$l_{min} < l_{i,j} < l_{max} \tag{4}$$

l_{min} and l_{max} are determined empirically by pre-learning.

2.3 Clustering Optical Flows

Next, in the proposed method, optical flows are clustered based on the lengths, angles, and source locations of the detected optical flows, as shown in Figure 7. Clustering optical flows is important for distinguishing multiple pedestrians who are walking together. In Section 3, we explain the detail of clustering optical flows.

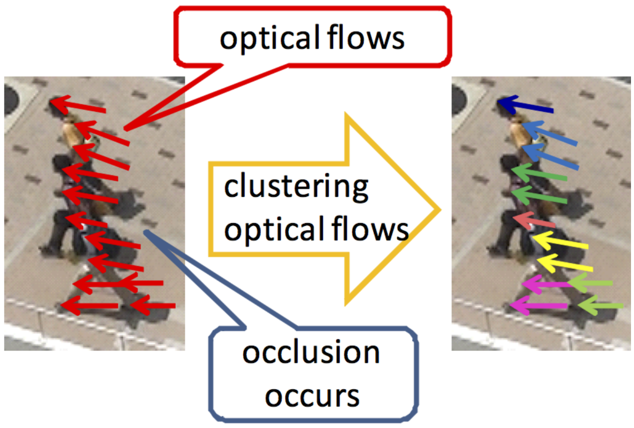


FIGURE 7: Optical Flow Clustering.

2.4 Estimating the Number of Pedestrians

We assume that the number of clusters detected in a frame and the actual number of the pedestrians in the frame have a strong correlation. Since optical flows are extracted from moving parts which indicate pedestrians in video frames, and optical flows detected from a single pedestrian tend to belong to the same cluster. The degree of correlation depends on the camera position, the size of pedestrians in the frames, and various other factors. The method utilizes a part of frames in the video sequence for pre-learning and estimates the number of pedestrians for all frames in the video sequence. Here, the number of clusters in frame f_i is denoted by C_i. The method calculates the average number of clusters in the frames for pre-leaning as follows:

$$\bar{C} = \frac{\sum_{q=1}^{F_l} C_q}{F_l} \tag{5}$$

where F_l $(< F)$is the number of frames for pre-learning. The actual number of pedestrians in frame f_i is denoted by P_i. The method obtains the average number of pedestrians in the frames for pre-learning as follows.

$$\bar{P} = \frac{\sum_{q=1}^{F_l} P_q}{F_l} \tag{6}$$

Note that to calculate $\bar{P}$, the method must count the actual number of pedestrians in the part of the video sequence, which means pre-learning.

The average number of clusters per person in pre-learned frames, denoted by $\bar{C_p}$, is calculated as follows.

$$\bar{C_p} = \frac{\bar{C}}{\bar{P}} \tag{7}$$

The estimated number of pedestrians in frame f_i, denoted by P_i^e, is calculated as

$$P_i^e = \frac{C_i}{\bar{C_p}} \tag{8}$$

3. OPTICAL FLOW CLUSTERING

It is assumed that optical flows detected from a single pedestrian have similar lengths, angles, and since their starting points exist within a certain range. Therefore, optical flows are clustered based on the degree of similarity of their lengths, angles, and starting points. Here, a set of detected optical flows in frame f_i is denoted by D_i and the l-th optical flow in D_i is denoted by $o_{i,l}$ $(\in D_i)$.

The proposed method conducts the clustering all optical flows in D_i. At first, a new cluster is created, denoted by $G_{i,1}$, and the optical flow $o_{i,1}$ $(\in D_i)$ is added to $G_{i,1}$. For the optical flow $o_{i,l}$ $(l \geq 2)$, $o_{i,l}$ is compared with all optical flows in each cluster $G_{i,m}$, in terms of length, angle, and starting point. We explain each step of optical flow clustering using two optical flows $o_{i,l}$ and $o_{i,s}$.

The lengths of two optical flows $o_{i,l}$ and $o_{i,s}$ are first compared where the following equation is satisfied or not:

$$|l_{i,l} - l_{i,s}| \leq l_{th} \tag{9}$$

where $l_{i,l}$ and $l_{i,s}$are the lengths of $o_{i,l}$ and $o_{i,s}$, which are represented by following equations (Eq.(10) and Eq.(11)), and l_{th} is the threshold of for clustering optical flows.

$$l_{i,l} = \sqrt{(x_{i,l}^e - x_{i,l}^s)^2 + (y_{i,l}^e - y_{i,l}^s)^2} \tag{10}$$

$$l_{i,s} = \sqrt{(x_{i,s}^e - x_{i,s}^s)^2 + (y_{i,s}^e - y_{i,s}^s)^2} \tag{11}$$

where $(x_{i,l}^s, y_{i,l}^s)$ and $(x_{i,l}^e, y_{i,l}^e)$ are the starting point and ending point of optical flow $o_{i,l}$, respectively.

Next, the angles of two optical flows $o_{i,l}$ and $o_{i,s}$ are next compared where the following equation is satisfied or not:

$$|\theta_{i,l} - \theta_{i,s}| \leq \theta_{th} \tag{12}$$

where $\theta_{i,l}$ and $\theta_{i,l}$ are the angles of $o_{i,l}$ and $o_{i,s}$, which are represented by following equations (Eq.(13) and Eq.(14)), and θ_{th} is the threshold of angle.

$$\theta_{i,l} = \arctan\left(\frac{y_{i,l}^e - y_{i,l}^s}{x_{i,l}^e - x_{i,l}^s}\right) \tag{13}$$

$$\theta_{i,s} = \arctan\left(\frac{y_{i,s}^e - y_{i,s}^s}{x_{i,s}^e - x_{i,s}^s}\right) \tag{14}$$

Finally, the starting points of two optical flows $o_{i,l}$ and $o_{i,s}$are compared.

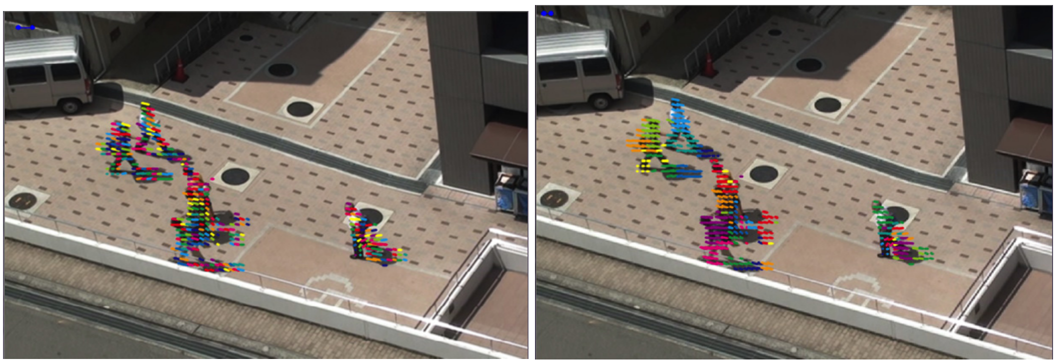
$$|x_{i,l}^s - x_{i,s}^s| \leq x_{th} \tag{15}$$

$$|y_{i,l}^s - y_{i,s}^s| \leq y_{th} \tag{16}$$

where x_{th} and y_{th} are respectively the thresholds of x-coordinate and y-coordinate of optical flows.

When $o_{i,l}$ and all optical flows in $G_{i,m}$ satisfy Eq.(9)–(16), the proposed method adds $o_{i,l}$ to $G_{i,m}$. When $o_{i,l}$ does notbelong to any existing cluster, a new cluster for $o_{i,l}$ is created.After clustering, the number of optical flows in each cluster is assessed. When the number is smaller than the threshold N_d, the cluster is discarded since we assume the cluster is caused by noise such as changes in the illumination conditions.

Figures 8(a) and 8(b) show the clustering results when θ_{th} is small and large, respectively. In these figures, the colored lines represent detected optical flows, and the optical flows that have the same color belong to the same cluster. When θ_{th} is small, since the number of optical flows in each cluster is small, the number of clusters is large. On the other hand, when θ_{th} is large, since the number of optical flows in each cluster is large, the number of clusters is small. Such difference with difference value of θ_{th} would affect the counting accuracy of the proposed method. In Subsection 4.3, the effect of θ_{th} on thecounting accuracy is evaluated.



(a) Small θ_{th} (b) Large θ_{th}

FIGURE 8: Results of Optical Flow Clustering.

4. EXPERIMENTAL EVALUATION

4.1 Experimental Environment

The effectiveness of the proposed method was evaluated through implementation experiments with three video sequences that have different characteristics. One was recorded a 30 [s] video (720×480 [pixels]) at the Toyonaka Campus of Osaka University, Japan at 29 [frames per second (fps)]. The shadows of pedestrians lay across the pavement in this video. Second video was recorded at a pedestrian crossing in Tokyo, Japan. In the video, many pedestrians move in different directions. The video (320×240 [pixels]) is 20 [s] long and was recorded at 29 [fps]. There is no shadow and the size of pedestrians is large in this video. The other is from the PETS contest, which was obtained from [40]. The PETS contest video (768×576 [pixels]) is 34 [s] long and was recorded at 7 [fps]. In this video, many pedestrians are present and occlusion occurs frequently. A snapshot from each video sequence is shown in Figures 9(a), 9(b), and 9(c), respectively. The proposed method was implemented on a PC with a 2.93 [GHz] Intel Core i7 CPU and 3.00 [GB] of memory, running Microsoft Windows 7 Enterprise. The parameters of the method for all video sequences are set as follows: α = 0.05, l_{min} = 0.2, l_{max} = 25.0, F_l = , θ_{th} = 0.2, l_{th} = 3.0, x_{th} = 38.0, y_{th} = 46.0, and N_d= 13. An example of the execution of the proposed method is shown in Figure 10.

In performance evaluations, we compared the counting accuracy of the proposed method with a non-clustering method that estimates the number of pedestrians based on a pre-learned correlation between the total number of optical flows, not clusters, and the actual number of pedestrians in each frame. The other parts of the non-clustering method are the same as in the proposed method.

(a) Osaka Univ. video (b) Tokyo video (c) PETS video

FIGURE 9: Snapshots of Video Sequences.

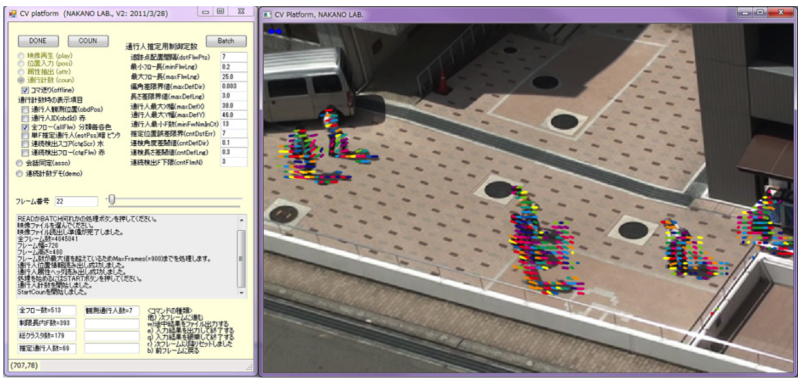
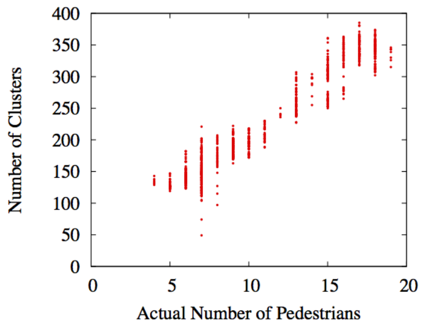


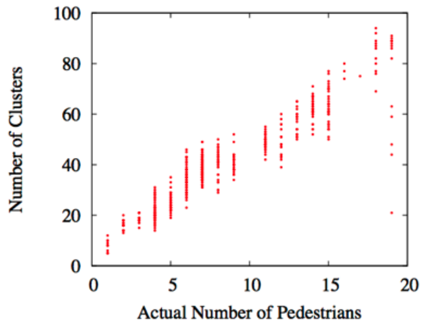
FIGURE 10: Snapshot of the Proposed Method.

4.2 Results and Discussions

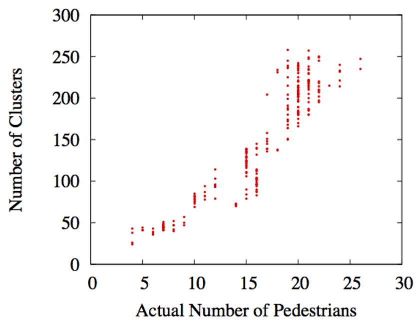
Figures 11(a), 11(b), and 11(c) show the relationship between the number of clusters detected by the proposed method and the actual number of pedestrians for three video sequences. Each plot shows the result for each frame. We observe a linear correlation between the number of clusters and the actual number of pedestrians regardless of the difference of the video sequence. This is because we utilize the simple pre-learning algorithm described in Subsection2.4. From figures 11(a), 11(b), and 11(c), the degrees of correlation between the number of clusters and the actual number of pedestrians are different, since the camera position, the size of pedestrians, and other factors are different. Therefore, the proposed method utilizes the correlation in the video sequence for pre-learning.



(a) Results of Osaka Univ. Video.



(b) Results of Tokyo Video.



(c) Results of PETS Video.

FIGURE 10: Correlations Between the Number of Clusters and the Actual Number of Pedestrians.

Tables 1(a), 1(b), and 1(c) summarize the evaluation results for each video sequence, which show the overall counting accuracy in terms of mean absolute error (MAE), mean relative error (MRE), and variance of the absolute error for the entire video sequence. For all video sequences, the proposed method outperforms the non-clustering method, indicating the effectiveness of the optical flow clustering proposed in this paper. For the video sequence taken in Tokyo in particular, the proposed method had counting accuracy that was about 25% higher than that of the non-clustering method. The proposed method had higher counting accuracy than that of the non-clustering method for another video sequences.

Method	MAE	MRE	Variance of absolute error
Proposed	0.952	0.103	1.211
Non-clustering	0.989	0.110	1.284

(a) Results of Osaka Univ. Video.

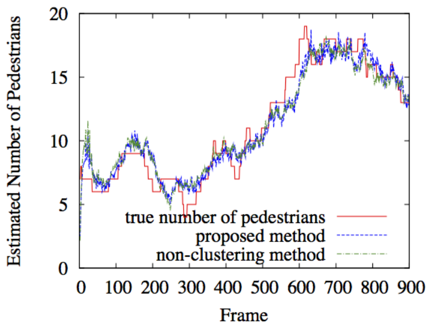
Method	MAE	MRE	Variance of absolute error
Proposed	1.123	0.170	1.603
Non-clustering	1.502	0.201	1.932

(b) Results of Tokyo Video.

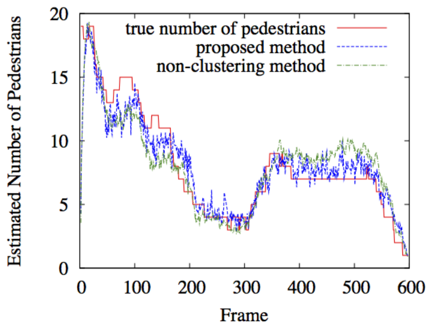
Method	MAE	MRE	Variance of absolute error
Proposed	2.528	0.174	3.164
Non-clustering	3.085	0.220	3.738

(c) Results of PETS Video.
TABLE 1: Overall Counting Accuracies.

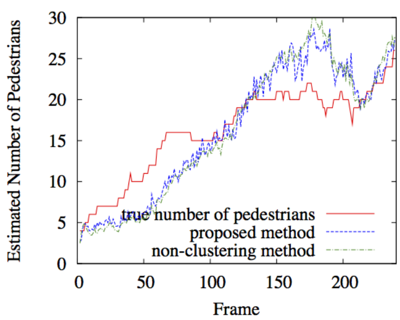
Figures 12(a), 12(b), and 12(c) show the temporal changes in the counting accuracy of the proposed method and that of the non-clustering method for each video sequence. The counting accuracy of both methods becomes worse especially when the actual number of pedestrians increases or decreases largely. One possible reason is that the method described in Subsection 2.2 fails to detect optical flows correctly when pedestrians enter or leave the target area. In such situations, a part of a pedestrian detected at the edge of the target area, which degrades the estimation accuracy based on pre-learning method. Other reasons for the lower counting accuracy include pedestrians that stop moving in the frame and the effect of shadow.



(a) Resuts of Osaka Univ. Video.



(b) Results of Tokyo Video.



(c) Results of PETS video.
FIGURE 12: Temporal Changes in Counting Accuracy.

It should be also noted that the counting accuracy of both methods significantly deteriorates for the video sequence from the PETS contest from Figure 12(c). In the PETS video sequence, tall and short pedestrians are walking together and the whole bodies of short pedestrians are sometimes completely occluded. Therefore, both methods cannot extract optical flows of occluded pedestrians. Additionally, since the frame rate of the video sequence from the PETS contest is extremely low, the pedestrians move largely between two successive frames. Therefore, the method described in Subsection 2.2 fails to detect optical flows.

Another reason for this large error is observed in Figure 11(c). Compared with Figures 11(a) and 11(b), the linear relationships between the number of pedestrians and the number of clusters is weak for PETS video sequence. Finding more appropriate fitting function is one of our future research.

4.3 Effect of Angle Threshold on Optical Flow Clustering

In this subsection, we focus on the angle threshold for optical flow clustering, since the angle threshold is important for distinguishing multiple pedestrians who are walking together. Figure 13 shows the relationship between the angle threshold and the MRE of the number of pedestrians estimated by the proposed method for the video sequence at Osaka University.

Note that the similar results have been obtained for the video sequence taken in Tokyo and that from the PETS contest. The finest angle resolution of the proposed method is approximately 0.2 degree, which is governed by the frame resolution, the average size, and the velocity of pedestrians. From this figure, it is found that the counting accuracy is high when the angle threshold is around 0.1 degree, which is roughly the same as the method’s upper limit of angle resolution. We recall Figures 8(a) and 8(b) for small and large angle thresholds, where the angle threshold in Figure 8(a) is 0.1 degree and that in Figure 8(b) is 5.0 degree. By comparing these figures, it is observed that the proposed method detects more clusters for each pedestrian when small threshold is utilized. However, by utilizing the linear correlation exhibited in Fig. 11(a), the proposed method can increase the counting accuracy.

Figure 14 shows the relationship between the actual number of pedestrians and the MRE of the number of pedestrians estimated by the proposed method, for various angle threshold values in video sequence at Osaka University. The error bars in the graph indicate the 95% confidence interval. We can see from this figure that for detecting a small number of pedestrians, a larger threshold is desirable, whereas a smaller threshold is preferable for detecting a larger number of pedestrians. These characteristics are caused by the effect of the pedestrians’ legs. When the number of pedestrians is small, less occlusion occurs and many optical flows are detected from the legs. Since the movement of the pedestrians’ legs is more complicated than that of other body parts, a large angle threshold is better for avoiding redundant and fluctuating segmentation of optical flow clusters. On the other hand, when there are many pedestrians, a substantial amount of occlusion can be observed and the legs of the pedestrian are often hidden. In such a case, we should utilize a small angle threshold to distinguish multiple pedestrians who move together causing occlusion. This is also confirmed by Figures 8(a) and 8(b).

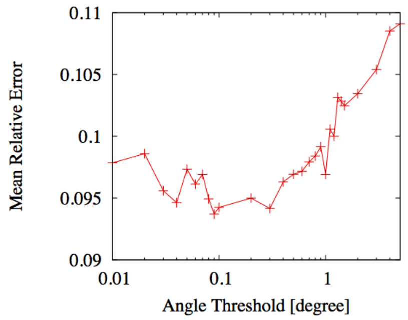


FIGURE 13: Effect of Angle Threshold on Counting Accuracy With Osaka Univ. Video.

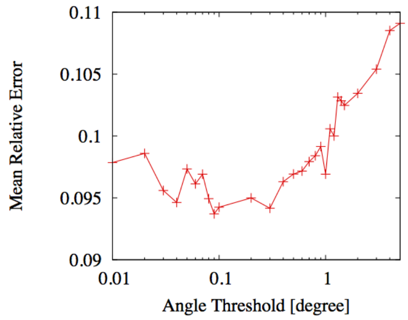


FIGURE 14: Effect of True Number of Pedestrians on Counting Accuracy.

4.4 Comparison of Two Algorithms of Extracting Optical Flows

In this subsection, we compared two algorithms of extracting optical flows, which utilize feature points and lattice points, explained in Subsection 2.2.

Figure 15 shows the relationship between the number of points set for extracting optical flows and the MRE of the number of pedestrians estimated by the proposed method using each algorithm in the video sequence at Osaka University. For both methods, the number of points is regulated by setting lower bound of pixels between two points. It is observed from this figure that the counting accuracy using feature points is higher than that using lattice points. The reason is that the algorithm using feature points detects the points almost from pedestrians' heads or shoulders. Since pedestrians' head and shoulders are hardly occluded even when severe occlusion occurs, the feature points are effective in extracting optical flows for distinguishing occluded pedestrians. On the other hand, the algorithm using lattice points utilizes the points equally separated. Therefore, the number of points utilized for extracting optical flows is only dependent on the size of moving part, not on the size of pedestrians included in the moving part. Then, the counting accuracy becomes worse especially when the number of occluded pedestrians increases.

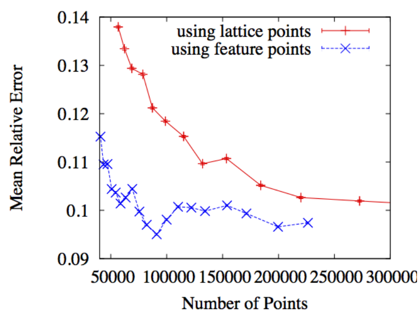


FIGURE 15: Performance Comparison of Two Algorithms for Extracting Optical Flows.

4 CONCLUSIONS

In this paper, we proposed an optical flow clustering method to improve the accuracy of counting pedestrians in the video sequences. In the proposed method, optical flows are clustered based on the lengths, angles, and source locations of optical flows. The proposed method counts the number of pedestrians using pre-learned statistics, based on the number of clusters detected by our method and the number of pedestrians having a strong and linear correlation. We evaluated the accuracy of the proposed method using several video sequences and showed that the counting accuracy using proposed method is up to 25% better than that of a non-clustering method. We also found that the clustering threshold of angles less than 1 degree is effective for enhancing counting accuracy. Additionally, in compared the performance of two algorithms which are use feature points and lattice points when optical flows are detected, the counting accuracy using feature points is higher than that using lattice points especially when the number of occluded pedestrians increases.

The most prioritized task on the present research work in the future is to compare the performance of the proposed method with other existing methods, to confirm the effectiveness of the proposed method. Especially, the comparison with method in latest PETS contest is important.

To emphasize wide applicability of the proposed method, we also plan to evaluate the proposed method using other video sequences to confirm the robustness of the proposed method. The improvement of the accuracy of the proposed method is necessary by resolving the problems explained in Subsections 4.2 and 4.3. By completing the above-mentioned tasks, we can completely confirm the advantage of the proposed method.

REFERENCES

[1] F. Bu, R. Greene-Roesel, M. Diogenes, and D. Ragland, “Estimating pedestrian accident exposure: Automated pedestrian counting devices report,” UC Berkeley Traffic Safety Center, Tech. Rep., Mar. 2007.

[2] X. Liu, P. H. Tu, J. Rittscher, A. G. A. Perera, and N. Krahnstoever, “Detecting and counting people in surveillance applications,” in Proc. AVSS 2005, pp. 306–311, Sep. 2005.

[3] H. Zhao and R. Shibasaki, “A novel system for tracking pedestrians using multiple single-row laser-range scanners,” Systems, Man and Cybernetics, vol. 35, pp. 283–291, Mar. 2005.

[4] A. Fod, A. Howard, and M. J. Mataric, “Laser-based people tracking,” in Proc. IEEE ICRA 2002, pp. 3024–3029, May 2002.

[5] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, “Laser-based detection and tracking of multiple people in crowds,” Computer Vision and Image Understanding, vol. 106, pp. 300–312, May 2007.

[6] C. Harris and M. Stephen, “A real-time system for monitoring pedestrians,” in Proc. WACV-MOTION 2005, vol. 1, pp. 378–385, Jan. 2005.

[7] A. Leykin and R. Hammoud, “Robust multi-pedestrian tracking in thermal-visible surveillance videos,” in Proc. Conference on CVPR Workshop 2006, p. 136, Jun. 2006.

[8] E. Goubet, J. Katz, and F. Porikli, “Pedestrian tracking using thermal infrared imaging,” in Proc. SPIE 2006, no. 6206, pp. 797–808, Apr. 2006.

[9] K. Hashimoto, M. Yoshinomoto, S. Matsueda, K. Morinaka, and N. Yoshiike, “Development of people-counting system with human-information sensor using multi-element pyroelectric infrared array detector,” Sensors and Actuators A: Physical, no. 58, pp. 165–171, Feb. 1998.

[10] SenSource Inc, “PCW-2BX03 Directional People Counter.” Internet: www.sensourceinc.com/PDF/PCW-Wired-Directional%20-%20Brochure.pdf. [Nov. 22, 2012].

[11] Eco counter, “People counters - Eco-counter.” Internet: www.eco-compteur.com. [Nov. 22, 2012].

[12] InfraRed Integrated Systems Ltd, “IRISYS people counter,” Internet: irisys.co.uk/people-counting. [Nov. 22, 2012].

[13] K. Hashimoto, M. Yoshinomoto, S. Matsueda, K. Morinaka, and N. Yoshiike, “Development of people-counting system with human-information sensor using multielement pyroelectric infrared array detector,” Sensors and Actuators, vol. 58, no. 2, pp. 165–171, Feb. 1997.

[14] Q. Chen, M. Gao, J. Ma, D. Zhang, L. M. Ni, and Y. Liu, “Mocus: Moving object counting using ultrasonic sensor networks,” International Journal of Sensor Networks, vol. 3, no. 1, pp. 55–65, Dec. 2007.

[15] T. Zhao and R. Nevatia, “Bayesian human segmentation in crowded situations,” in Proc. CVPR 2003, pp. 459-466, Jun. 2003.

[16] D. Conte, P. Foggia, G. Percannella, and M. Vento, “Performance evaluation of a people tracking system on pets2009 database,” in Proc. ACSS 2010, pp. 119–126, Aug. 2010.

[17] P. Kilambi, E. Ribnick, A. J. Joshi, O. Masoud, and N. Papanikolopoulos, “Estimating pedestrian counts in groups,” Computer Vision and Image Understanding, vol. 110, pp. 43-59, Apr. 2008.

[18] James Ferryman, “PETS 2010.” Internet: www.cvg.rdg.ac.uk/PETS2010. [Nov. 22, 2012].

[19] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Transactions on Neural Networks, vol. 24, no. 7, pp. 971–989, Jul. 2002.

[20] S. Bileschi and L. Wolf, “A unified system for object detection, texture recognition, and context analysis based on the standard model feature set,” in Proc. BMVC 2005, Sep. 2005.

[21] M. Heeikkila and M. Pietikainen, “A texture-based method for modeling the background and detection moving objects,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 657–662, Apr. 2006.

[22] V. Mahadevan and N. Vasconcelos, “Background subtraction in highly dynamic scenes,” in Proc. CVPR 2008, pp. 1–6, Jun. 2008.

[23] T. Zhao and B. Wu, “Segmentation and tracking of multiple humans in crowded environments,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 7, pp. 1198–1211, Jul. 2008.

[24] H. Kitazawa, Z. Li, and K. Yabuta, “Moving object extraction and tracking based on the exclusive block matching,” IEICE technical report, vol. 108, no. 4, pp. 49–54, Apr. 2008.

[25] N. Baba, T. Ejima, and H. Matsuo, “Headfinder: A real-time robust head detection and tracking system,” in Proc. SPIE Electronic Imaging Conference on Real-Time Imaging 2002, vol. 4666, pp. 42–51, Feb. 2002.

[26] A. Ess, B. Leibe, K. Schindler, and L. V. Gool, “A mobile vision system for robust multi-person tracking,” in Proc. CVPR 2008, pp. 1–8, Jun. 2008.

[27] R. Eshel and Y. Moses, “Homography based multiple camera detection and tracking of people in a dense crowd,” in Proc. CVPR 2008, pp. 1–8, Jun. 2008.

[28] T. Teixeira and A. Savvide, “Lightweight people counting and localizing in indoor spaces using camera sensor nodes,” in Proc. ICDSC 2007, pp. 36–43, Sep. 2007.

[29] V. Kettnaker and R. Zabih, “Counting people from multiple cameras,” in Proc. MCS 1999, vol. 2, pp. 267–271, Jul. 1999.

[30] A. Mittal and L. S. Davis, “M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo,” Computer Vision, vol. 2350, pp. 189–203, May 2002.

[31] J. P. Batista, “Tracking pedestrians under occlusion using multiple cameras,” in Proc. ICIAR 2004, vol. 3212, pp. 555–562, Oct. 2004.

[32] N. Friedman and S. Russell, “Image segmentation in video sequences: A probabilistic approach,” in Proc. UAI 1997, Aug. 1997.

[33] A. Elegammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” in Proc. ECCV 2000, vol. 2, pp. 751–767, Jun. 2000.

[34] T. Horprasert, D. Harwood, and L. Davis, “A statistical approach for real-time robust background subtraction and shadow detection,” in Proc. IEEE ECCV 1999 Frame-Rate Workshop, Sep. 1999.

[35] J.-S. Hu, T.-M. Su, and S.-C. Jeng, “Robust background subtraction with shadow and highlight removal for indoor surveillance,” in Proc. IROS 2006, pp. 4545–4550, Oct. 2006.

[36] N. Friedman and S. Russell, “Image segmentation in video sequences: A probabilistic approach,” in Proc. UAI 1997, pp. 175–181, Aug. 1997.

[37] C. Harris and M. Stephen, “A combined corner and edge detector,” in Proc. AVC 1988, pp. 147–152, Aug. 1988.

[38] R. Mehrotra and S. Nichani, “Corner detection,” Pattern Recognition, vol. 23, pp. 1223–1233, Mar. 1990.

[39] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in Proc. the 1981 DARPA Image Understanding Workshop, pp. 121–130, Apr.1981.

[40] James Ferryman, “Video sequence take from PETS workshop,” available Internet: www.cvg.rdg.ac.uk/PETS2010/a.html. [Nov. 22, 2012].