A PEDESTRIAN COUNTING SCHEME FOR VIDEO IMAGES

Chi-Cheng Cheng and Yi-Fan Wu

Department of Mechanical and Electro-Mechanical Engineering,
National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C

ABSTRACT

Pedestrian counting aims to compute the numbers of pedestrians entering and leaving an area of interest based on object detection and tracking techniques. This paper proposes a simple and effective approach of pedestrian counting that can effectively solve the problem of pedestrian occlusion.Firstly, the moving objects are detected by the median filtering and foreground extraction with the improved mixed Gaussian model. And then the HOG (Histogram of oriented gradient) features detection and the SVM (Support vector machine) classification are applied to identify the pedestrians. A pedestrian dataset containing 1500 positive samples, 12000 negative samples, and 420 hard examples, which gave the false discriminant results with the initial classifier, also considered as negative samples to enhance classification capability is employed. In addition, the Kalman filtering with BLOB analysis for dynamic target tracking is chosen to predict pedestrian trajectory.This method greatly reduces the target misjudgment caused by overlapping and completes the two-way counting. Experiments on pedestrian tracking and counting in video images demonstrate promising performance with satisfactory recognition rate and processing time.

KEYWORDS

Machine Vision, Kalman Filtering, Pedestrian Identification, Target Tracking, Pedestrian Counting.

1. INTRODUCTION

Pedestrian identification and people counting recently gather attentions of researchers and engineers in the fields of machine vision and intelligent security monitoring. These technologies have been useful for marketing analysis and security purposes and can be found in theatres, markets, malls, department stores, exhibition halls, transportation stations, government buildings, personnel controlled laboratories and many other places. Although pedestrian counting can be accomplished by using inferred or laser technologies, machine vision possesses advantages of broad spatial coverage, accurate results, cost effectivity, and ability to provide much more information about pedestrians by extracting features of people. Therefore, this paper aims to develop an effective and efficient framework for pedestrian identification and bidirectional counting for potential future applications.

Basically, there are two difference approaches to deal with the challenge of people counting. The first approach performs people counting based on moving regions where pedestrian is treated as a single moving object [1]. If the size of a moving region is similar to that of a pedestrian, the region will be counted as a person. However, if the size of a moving region is greater than that of a single person, the moving region needs to be decomposed into a number of areas for single

David C. Wyld et al. (Eds): ICAIT, CBIoT, WiMo, CRYPIS, ICDIPV, CAIML, NLCA - 2022
pp. 135-149, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121211

person according to prior knowledge regarding the pedestrian size. These methods highly rely on prior knowledge and bring about characteristics of low accuracy and fast computation. Another approach is based on image features and machine learning techniques. People counting is therefore achieved by template matching through learning process with image samples of pedestrian [2]. Although these methods provide much better performance in terms of accuracy, expensively computational cost and requirement of a large amount of image samples are their drawbacks.

Methods for pedestrian identification can be classified into four categories. The human body model approach applies geometric features of human body to justify if the region of interest is pedestrian or not. Most common features for identification of human body include head [3], trunk [4], hair [5], shoulders [6], etc. Furthermore, a straight ellipse was suggested to model the outline of the human's body [7].

The template matching approach searches for similar objects in images to identify pedestrian according to given templates. A grid-based template matching algorithm was introduced for people counting [8]. Besides, the fast Hough transform was applied to quickly search for potential position of human's head in the foreground [9]. Nevertheless, human's features cannot be always described by simple mathematical rules and extensive learning through lots of image samples is strongly required. The outline classification approach conducts object identification based on trained templates through a learning process with lots of image samples. A multilevel HOG-LBP (Histogram of oriented gradient-Local binary pattern) scheme based on the PCA (principal component analysis) was proposed to identify the profile of human's head and shoulders [10]. A boosting learning algorithm identifying human's head, shoulders, trunk, and legs was presented for detection and tracking of humans [11]. In addition, a people counting system based on detection and tracking of human face with a neural network training process was proposed [12]. The LDA (linear discriminant analysis) was chosen to enhance learning capability to human's features [13].

The movement feature approach is to perform pedestrian identification and tracking based on periodic movement characteristics of humans. Human movements can be assumed to be independent events. As a result, similar moving patterns should belong to the same human. Pedestrian identification was achieved by tracking corners with similar moving patterns followed by classification with a Bayesian framework [14].

2. EXTRACTION OF MOVING TARGETS

The purpose of pedestrian counting is to calculate the numbers of people entering and leaving a given region. In order to conduct pedestrian counting, pedestrian identification and moving objects tracking are usually involved. This paper proposes a systematic approach including foreground extraction, objects detection, objects tracking, and pedestrian counting.

Because image quality is strongly affected by weather, illumination, electromagnetic interferences, and other possible noises. It is required to include image pre-processing to remove unnecessary signals and enhance image quality at the beginning stage. The median filter is selected for image preprocessing because of its outstanding performance of noise removal and edge reservation.

2.1. Foreground extraction

Foreground extraction is to isolate regions with pedestrians in the images for the purpose of improving computational efficiency by narrowing searching area. There exist a number of popular methods for foreground extraction such as temporal difference, background subtraction, the optical flow. The temporal difference locates the moving objects relying on the difference between two or three successive images. This method is highly based on the assumption of same background contents and can only be applied to images taken under invariant illumination environment [9, 15]. The background subtraction employs the difference between the image and the known background to achieve moving objects detection. Nevertheless, a reliable background image, which can adapt to variant background, plays a crucial role to successfully extract moving objects from the background [10,16]. The optical flow method extracts moving objects according to the derivatives of brightness stemming from the constraint of brightness consistency Although this approach demonstrates satisfactory detection performance for moving objects, static objects cannot be successfully identified [11,14].

In order to extract moving and static pedestrian in video images, a foreground extraction scheme based on template matching is therefore demanded. A hierarchical template matching algorithm using contour features was developed due to its significant detection performance and computational efficiency. Lots of samples as templates are required to reflect many possible postures of pedestrian. Nevertheless, reduction of computation speed will be resulted. As a result, a combined coarse-to-fine approach in shape and parameter space was proposed to enhance computational efficiency [12,17].

The single Gaussian model, which consists of background initialization and background update, is quite applicable to interior environment and uncomplicated outdoor space. However, the single Gaussian model may fail due to non-constant illumination, varying environment, and incursion and vanishing of unknown objects. Consequently, the mixture-of-Gaussians classification model was introduced to deal with both robustness to environment and real-time capability by Stauffer and Grimson [20,18]. Three stages, establishment of background model, identification of background model, and update of background model, are involved in this technique for foreground extraction.

At the establishment of background model stage, three important parameters regarding the Gaussian distribution including mean, standard deviation, and its weighting factor, are required to be determined. As for the identification of background model, assume there are K Gaussian distributions with corresponding weighting factors w_k. a priority list can be arranged according the ratio of the weighting factor to the standard deviation, w_k/σ_k. The weighting factor and the standard deviation indicate the duration time and the stability of the Gaussian distribution, respectively. If the sum of the weighting factors for the top B Gaussian distributions in the list is larger than a given threshold value T, these B distributions can be applied as the background model and the rest distributions will represent the foreground.

After the background model is established, if pixels of a new coming video sequence agree with the model, they will be classified as the background, otherwise foreground. A standard process to justify if a pixel belongs to the background can be formulated by

$$|x_t - \mu_{t-1}| < 2.5\sigma_{t-1}$$

where x_t is the pixel information at time t, and μ_{t-1} and σ_{t-1} stand for the mean value and the standard deviation of the Gaussian distribution at time t-1, respectively. The background model can therefore be updated by modifying corresponding parameters for each Gaussian distribution.

Unfortunately, this approach may suffer from update failure due to slow learning rate. Consequently, a novel learning rate formula and an online expectation-maximization (EM) algorithm were proposed to improve convergent speed and adaptation capability to variant environment [21,19].

It was found that the mixture-of-Gaussians classification scheme was able to provide complete contours of moving objects in the foreground. However, the result also contains partial contour on the background with noises especially for situation with complicated background. As a result, the following modified online EM algorithm is proposed to maintain complete contour information of moving objects in the background and eliminate fake foreground pixels as well as background noises.

1. Initially, apply the original online EM algorithm to obtain the background image I_b and the binarized foreground image I_f.
2. Find the difference image I_d by subtracting I_b from the current image I.
3. Scan every pixel in the image I_f. If the pixel belongs to the background, leave it unchanged; otherwise, fill it with the pixel information at the corresponding location of I_d.
4. Implementation of a thresholding process to the modified foreground image I_f.
5. Calculate boundary lengths of contours and the area size enclosed by contours. If either the boundary length or the area size is too small, remove the correspond contour.
6. The foreground image is updated by filling the contour with the foreground intensity.

Remarkable performance of foreground extraction on sequences of images taken from the Intelligent Room video in SBMnet (http://scenebackgroundmodeling.net) dataset and the PETS (Performance evaluation of tracking and surveillance) 2009 dataset is illustrated in Figure 1.

Figure 1. Foreground extraction using the proposed modified EM algorithm.

2.2. Objects detection

Objects detection is to identify pedestrian from the foreground as the targets for objects tracking afterwards. After the process of foreground extraction, moving objects are successfully identified. Nevertheless, moving objects may consist of pedestrian and other matters with movement.

Basically, there are four different approaches to deal with objects detection. They are model-based, template matching, posture classification, and movement-based methods. There have been a number of feature detection methods for pedestrian identification including the Harr-like technique, local binary patterns (LBP), histogram of oriented gradient (HOG), and pyramid histogram of oriented gradients (PHOG). In order to accommodate computational efficiency and detection performance, the approach of HOG is chosen for this research.

2.2.1. HOG

The approach of the histograms of oriented gradients locates features of local regions in an image.The followings are simplified HOG procedures:

1. Set the size for the window, which stands for the region of features extraction.
2. Compute the amplitude and the angle for the gradients in the window.
3. Partition the window into a number of overlapped blocks and decompose a block into un-overlapped cells. Establish the histogram for the angles of gradients in each cell.
4. Combine all histograms in a block to form a histogram vector.
5. A complete HOG features vector is constructed by collecting histogram vectors for all blocks

After the features of an image are extracted by the histograms of gradients, the technique of the support vector machine (SVM) will be applied to determine a best hyperspace as the decision function for classification.

2.2.2. SVM

The SVM aims to mapping feature vectors to a hyperspace so that a best hyperplane can be found for classification. Assume $\{-1,+1\}$ represents two different classes. Training samples are denoted by $\mathbf{x}_i$ and its corresponding output is $y_i \in \{-1,+1\}$. Since there are only two possible classes, the decision function for classification can be formulated by $\mathbf{w}^T\mathbf{x} + b = 0$. In order to allow the data can be classified by the maximum-margin hyperplane, i.e., the distance between the hyperplane and the nearest point $\mathbf{x}_i$ from either group is maximized, the constraint equation can therefore be defined as

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$$

Consequently, the desired function becomes

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^T\mathbf{x} + b)$$

where $\mathbf{x}$ is the vector of sampled data and $\mathbf{w}$ is the normal vector to the hyperplane.

If the training data is separable, two parallel hyperplanes can be chosen to divide the data into two classes. Geometrically, the distance between these two hyperplanes can be found to be $2/\|\mathbf{w}\|$.Enlarge the distance of those two hyperplanes is actually equivalent to minimize $\|\mathbf{w}\|$. Therefore, this optimization problem can be simply expressed by

$$\min_{\mathbf{w},b} \frac{\|\mathbf{w}\|}{2} \text{ subject to} y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$$

The optimization problem listed above can be solved using the Lagrange multiplier approach by reformulating the problem as

$$\frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i \{y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\}$$

where α_i is the Lagrange multiplier and *N* denotes the total number of data point. Since the minimal needs to be reached, the partial derivatives with respect to **w** and *b* must be zero and the following dual problem can be obtained.

$$\min_{\alpha_i} \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \text{ subject to } \sum_{i=1}^{N} \alpha_i y_i = 0$$

After solving α_i and b, for any given vector *x* that needs to be classified, the object function becomes

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w}^T \text{x} - b) = \text{sgn}\{\sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} - b\}$$

2.2.3. Pedestrian Identification based on HOG and SVM

The process of pedestrian identification based on HOG and SVM is illustrated as Figure 2. The HOG is applied for features extraction and followed by the SVM classifier for pedestrian identification.

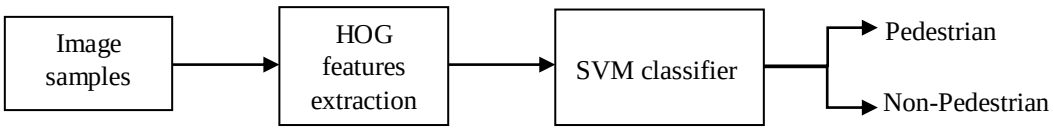

Figure 2. Flowchart of pedestrian identification based on HOG and SVM.

The positive pedestrian samples are from the INRIA pedestrian dataset, which contains 2416 pedestrian images with resolution of 64x128. Basically, the peoples in this dataset are all in standing posture, but with different appearance and clothing. Figure 3 depicts some image examples from the INRIA pedestrian dataset. There are 1218 negative pedestrian sample imageswith different sizes in the dataset. A set of 12180 negative images in total was collected by randomly cropping 10 64x128 regions from each sample. A number of negative pedestrian image samples are shown as in Figure 4. Through careful examination on those sample images, some of them, not appropriate for positive samples, were moved to the set of negative samples or even removed from the training data to enhance identification capability of the SVM classifier. Consequently, there are 1500 positive samples and 12000 negative samples in total in the training process.

Figure 3. Some positive pedestrian image samples from the INRIA dataset.

Figure 4. Some negative pedestrian image samples.

3. PEDESTRIAN TRACKING AND COUNTING

After successful foreground extraction and identification of pedestrian, pedestrian tracking and counting will be implemented by combining techniques of the Kalman filtering and BLOB (Binary large objects) analysis.

3.1. Kalman filtering [18,20]

The Kalman filtering has been an effective computational approach for tracking of a moving object. It provides a systematic recursive algorithm according to a state-space dynamic equation and an observation model including possible state estimation errors $\mathbf{w}_k$ and measurement noises $\mathbf{v}_k$, i.e.,

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k$$
$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ can be modelled as normal statistical distributions of $N(0, \mathbf{Q}_k)$ and $N(0, \mathbf{R}_k)$, respectively. $\mathbf{Q}_k$ and $\mathbf{R}_k$ denote the covariance matrices of the process noise and the observation noise.

The whole Kalman filtering scheme consisting of a prediction stage and an update stage can be illustrated as in Figure. 5. The prediction stage estimates the system's states $\hat{\mathbf{x}}_k^-$ and the covariance of the predicted error $\mathbf{P}_k^-$ before the measurement $\mathbf{z}_k$. The corresponding predicted error e_k^- is therefore defined as $\mathbf{x}_k - \hat{\mathbf{x}}_k^-$. At the update stage, the system's state and the covariance matrix of the predicted error are modified as $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ respectively when the new measurement z_k is received. $\mathbf{K}_k$ is known as the optimal Kalman gain. Both the prediction stage and the update stage are recursively executed to minimize the covariance matrix of the predicted error.

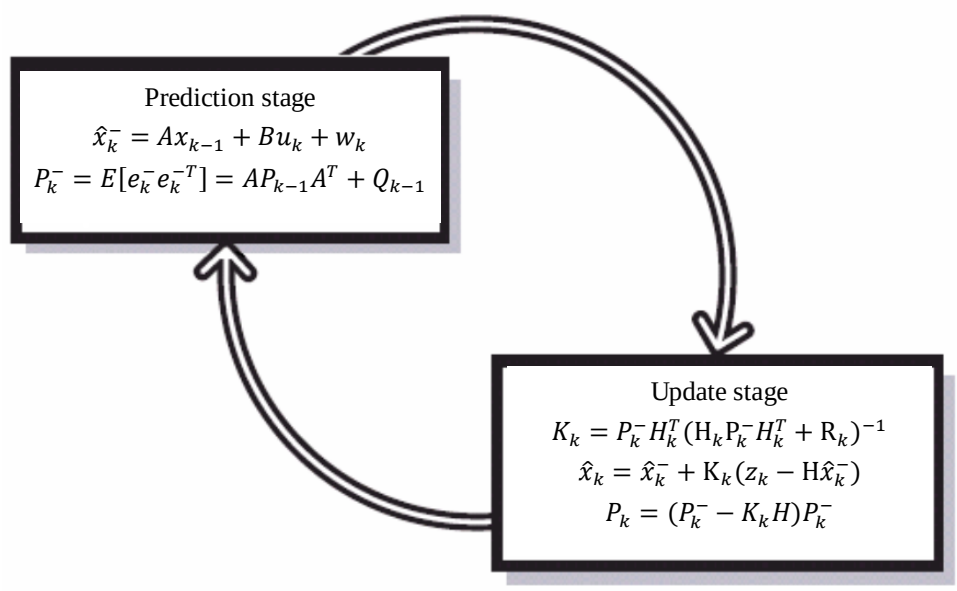

Figure 5. Two stages of the recursive Kalman filtering algorithm.

3.2. BLOB analysis

After the process of object detection, a sequence of foreground and background images are obtained. The foreground images can be represented by black-and-white binary images using the standard binarization method. As a result, the foreground consists of several connected regions. In order to acquire features of the foreground, some geometric properties such as location, size, perimeter, boundary length, and moment of inertia need to be calculated. Location and size of a connected region can be determined by its geometric center and the sum of number of pixels inside the region. The spatial moments can be applied for describing geometric characteristics of an image region because of its rotational invariant property. The general expression for the spatial moments $m_{p,q}$ can be written as

$$m_{p,q} = \sum B(x,y)x^p y^q$$

where $B(x,y)$ is the binary value either 1 or 0 at (x,y) in the image plane, and p and q stand for the order of the moment with respect to the x and y dimensions, respectively. If both p and q are all zeros, $m_{0,0}$ actually indicates the area in that image region.

The BLOB matching method takes advantage of image features such as shape, size, and the spatial moments to search for the target object and belongs to a bottom-to-top tracking approach. Apparently, this method is quite effective for rigid objects and limited number of moving objects. If the number of moving objects is large, it will be computationally expansive for the matching process. Nevertheless, the Kalman filtering technique predicts the moving object's position based on previous motion information of the moving object and is classified as a top-to-bottom tracking method. Since the Kalman filtering technique is able to predict the future position of the moving object, the BLOB matching algorithm is only applied to the nearby of the estimated location so that computational efficiency can be greatly improved. Therefore, pedestrian tracking and counting proposed in this paper will be achieved by a hybrid scheme combining both Kalman filtering and BLOB matching techniques.

Important procedures for pedestrian tracking are summarized as follows:

1. According the binary foreground generated by the improved Gaussian approach, search for connected regions. Mark each connected region as a BLOB and calculate its geometric center, contour length, and spatial moments as region features.
2. Based on current geometric centers of all BLOBs, estimate their locations for the next frame using the Kalman filtering method by defining

$$\mathbf{x}_k = [x, y, w, h, \dot{x}, \dot{y}, \dot{w}, \dot{h}]^T$$
$$\mathbf{z}_k = [x, y, w, h]^T$$

where x and y represent the coordinate of the geometric center, and w and h are the width and length of the BLOB. In addition, the control-input matrix **B** is a null matrix, and the state transition matrix **A** and the observation matrix **H** are respectively given as

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

3. Search for the matched BLOB with similar contour length and spatial moments near the estimated locations.

3.3. Pedestrian counting

Pedestrian counting is accomplished by computing the net number of pedestrian crossing a given screen line. Assume the origin of the coordinate system for the image is located at the upper left corner based on the conventional setup of the image reference frame and two end points of the screen line are defined as $P_1(a_1, b_1)$and $P_2(a_2, b_2)$. The n-th pixel along the screen line can be written as

$$P(n) = (a_1 + ncos\theta, b_1 + nsin\theta)$$

where

$$\theta = \tan^{-1}\frac{b_2 - b_1}{a_2 - a_1}$$

Let $F_r(k)$ denote the foreground image at a sampling instant. If there is a moving object on the left side of the screen line, its corresponding region can be expressed by

$$P_{fr} = \{F_r(i,j,k) | F_r(i,j,k) = 255, i_1 < i < i_2, j_1 < j < j_2\}$$

where i_1, i_2, j_1, and j_2 respectively stand for the top, bottom, left, and right limits of the region for the moving object. Once every pixel in $F_r(i,j,k)$ satisfies

$$i > a_1 + \frac{j - b_1}{\sin\theta} \cdot \cos\theta$$

It can be concluded that the moving object has successfully crossed over the screen line from left to right. Unfortunately, it would be quite complicated to judge the traveling direction of pedestrian if more than one pedestrian is involved. In order to overcome this difficulty, a screen line is extended to a screen strip with a certain width bounded by two parallel dashed line as shown in Figure 6. The vertical distance between those two dashed lines $2d$ is set to a little bit wider than the width of a normal people. In other words, those two dashed lines can be represented by

$$P(n) = (a_1 + ncos\theta \pm d, b_1 + nsin\theta)$$

When a pedestrian passes the left dashed border line of the screen zone, the current frame number will be recorded and accumulation of pixels on the dashed line for forthcoming frames will be conducted. When the sum maintains unchanged, there are two possible conditions. If the moving region is located at the left-hand side of the left dashed line, the pedestrian's moving direction is from the right to the left. However, if the moving region is within the left dashed line and the screen line, the moving direction of the pedestrian should be from left to right. Similar process can be applied to determine the moving direction of the pedestrian passes the right dashed border line. This approach solves the difficulty on determination of moving directions for pedestrians passing over the screen line from both sides at the same time.

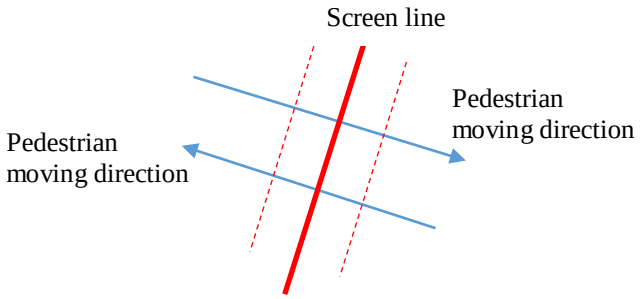

Figure 6. The screen line and two dashed border lines for pedestrian counting.

4. EXPERIMENTS

Experiments for pedestrian detection and counting were conducted by two prerecorded test videos from the internet and two live videos taken by authors. Software platform is the Visual Studio 2010 Ultimate with C++ programming language assisted by OpenCV 2.4.11 computer vision library. In addition, XviD codec was also applied for the purpose of video decoding. Before an experiment starts, a region of interest (ROI) with green line segments and a magenta screen line for pedestrian counting need to be given by users.

Two prerecorded test videos taken by monitoring cameras on the street were chosen for experiments and performance evaluation. Both test videos own 360P resolution. Test video #1 is with steady and little pedestrian flow without significant overlapping. Nevertheless, the pedestrian flow in test video #2 is much denser than that in test video #1. Besides, overlapped pedestrian frequently happens in test video #2. Figures 7 and 8 depict sample images in test video #1 and #2, respectively.

Live videos were taken by a Pantex digital single-lens reflex camera K-30 with 16.3-megapixel resolution. In order to have a sufficient height with an appropriate inclination angle downwards,

the camera was attached on a tripod standing on a table as illustrated in Figure 9. This setup was located on a hallway in campus of National Sun Yat-sen University. The height of the camera above the ground was 3.5 m and its depression angle below the horizontal level was 15 degrees so that the camera is able to capture the whole scene of the hallway. The live videos were set to 480P resolution. Sample images in live video #1 and #2 are depicted in Figures 10 and 11, respectively.

In order to evaluate the performance of pedestrian identification and counting, the accuracy η is defined as

$$\eta = 1 - \frac{|n_1 - n_2|}{n_1}$$

where n_1 and n_2 are actual and detected numbers of pedestrian, respectively. The average processing time is calculated based on the mean of computation time for five peoples by random selection.

Performance of pedestrian identification and counting for test videos is summarized in Table 1. Accuracy is a little bit reduced for test video #2 because of significant pedestrian overlaps. However, the proposed algorithm still demonstrates satisfactory performance in terms of accuracy. Without surprisingly, average processing time for test video #2 is greater than that for test video#1 due to more pedestrian involved in test video #2. A couple of errors were found in initial evaluation of pedestrian identification and counting for both live videos. The pedestrian images caused errors were therefore put into the group of negative samples for re-training. There exists two-way traveling direction for pedestrian in live video #2. The presented identification and counting strategy displays outstanding performance as shown in Table 2. Larger average processing time for live videos was mainly caused by better image resolution from 360P to 480P.

Figure 7. Sample images in test video #1.

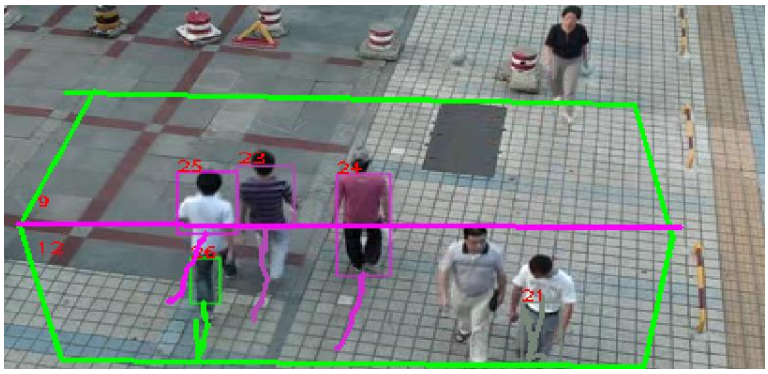

Figure 8. Sample image in test video #2.

Figure 9. Camera setup for experiments on pedestrian identification and counting.

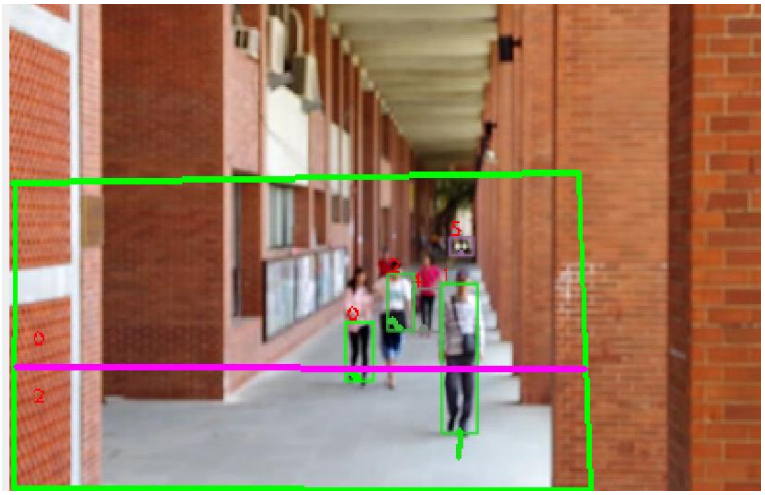

Figure 10. Sample image in live video #1.

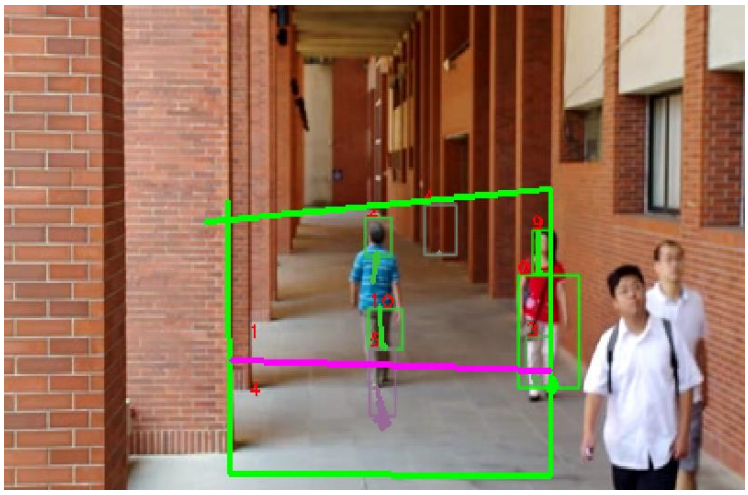

Figure 11. Sample image in live video #2.

Table 1. Performance of pedestrian identification and counting for test videos.

	Test video #1		**Test video#2**	
	Actual	Identified	Actual	Identified
Number of people entering	9	9	21	20
Number of people leaving	1	1	13	9
Accuracy	100%		82.8%	
Average processing time (ms)	35		60	

Table 2. Performance of pedestrian identification and counting for live videos.

	Live video #1		**Live video#2**	
	Actual	Identified	Actual	Identified
Number of people entering	6	6	5	5
Number of people leaving	0	0	1	1
Accuracy	100%		100%	
Average processing time (ms)	80		100	

5. CONCLUSIONS

This paper presents an effective and promising pedestrian counting scheme, which can be applied to areas required for control of people flow for the purposes of either security or marketing analysis. The proposed scheme consists of foreground extraction, pedestrian identification, pedestrian tracking, and counting of people flow. Foreground extraction is achieved by the improved mixed Gaussian model. Incorporating the HOG features detection with the SVM

classification is chosen for pedestrian identification. In order to enhance classification performance, those positive samples providing false discriminant results in the first classification run are moved to the group of negative samples. Furthermore, the Kalman filtering with BLOB analysis is employed to conduct dynamic target tracking for pedestrian trajectory prediction. Experiments on pedestrian tracking and counting for both dataset videos and live videos taken in campus environment demonstrate encouraging performance in terms of recognition rate and processing time. Above all, target misjudgment caused by overlapping can be greatly reduced and two-way counting becomes possible.

Nevertheless, further examinations on occlusion analysis and comparison with existing algorithms are required in the future work. Besides, in order to maintain portability for experimental setup, a notebook computer P770ZM with Intel® Xeon® E3-1231 v3 4x3.4 GHz was applied. If a more powerful desktop computer is chosen, real-time performance for actual applications can therefore be expected.

REFERENCES

[1] Li, F.S., Zhang, Y.C., Yang, H.C. & Wang, Y.P. (2014) "Fast pedestrians counting algorithm based on HOG", *Computer Systems & Applications*. Vol. 23, No. 5, pp. 172-176.

[2] Hsieh, J.W., Peng, C.S. & Fan, K.C. (2007) "Grid-based template matching for people counting",*Proc. IEEE 9th Workshop on Multimedia Signal Processing*, pp. 316-319.

[3] Vieren, C., Cabestaing, F. & Postarie, J.G. (1995) "Catching moving objects with snakes for motion tracking", *Pattern Recognition Letters*, Vol.16, pp. 679-685.

[4] Paviovic, V., Rehg, J., Cham, T.J. & Murphy, K. (1999) "A dynamic Bayesian network approach to figure tracking using learned dynamics models", *Proc. 7th IEEE Int. Conf. on Computer Vision*. Vol. 1, pp. 94-101.

[5] Zhao, M. (2008) "Hair-color modeling and head detection", *Proc. 7th World Congress on Intelligent Control and Automation*, pp. 7769-7772.

[6] Li, M., Zhang, Z.X. & Huang, K.Q. (2009) "Rapid and robust human detection and tracking based on omega-shape features", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2545-2548

[7] Zhao, T. & Nevatia, R. (2004) "Tracking multiple humans in complex situations", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, pp. 1208-1221.

[8] Hsieh, J.W., Peng, C.S. & Fan, K.C. (2007) "Grid-based template matching for people counting", *Proc. IEEE 9th Workshop on Multimedia Signal Processing*, pp. 316-319.

[9] Yuk, J.S.C., Wong, K-Y.K., Chung, R.H.Y., Chin, F.Y.L. & Chow, K.P. (2006) "Real-time multiple head shape detection and tracking system with decentralized trackers", *Proc. 6th International Conf. Systems Design and Applications*, pp. 384-389.

[10] Zeng, C. & Ma, H. (2010) "Robust head-shoulder detection by PCA-based multilevel HOG-LBP detector for people counting", *Proc. Int. Conf. Pattern Recognition*, pp. 2069-2072.

[11] Wu, B. & Nevatia, R. (2007) "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors", *Int. J. Computer Vision*, Vol. 75, No. 2, pp. 247-266.

[12] Zhao, X., Dellandréa, E. & Chen, L. (2009) "A people counting system based on face detection and tracking in a video", *Proc. Int. Conf. Advanced Video and Signal Based Surveillance*, pp. 67-72.

[13] Yu, R (2014) "Mobile app connecting people based on personality detection and image perception analysis". *Proc. IEEE Int. Sym. Multimedia*, pp. 330-340.

[14] Brostow, G.J. & Cipolla, R. (2006) "Unsupervised Bayesian detection of independent motion in crowds", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 594-601

[15] Collins, R. (2000) *A System for Video Surveillance and Monitoring: VSAM Final Report*. Technical Report CMU-RI-TR-00-12, Carnegie Mellon University.

[16] Barnich, O. & Droogenbroeck, M.V. (2011) "ViBe: A universal background subtraction algorithm for video sequences", *IEEE Trans. Image Processin*g, Vol. 20, pp.

[17] Gavrila, D. (2003) "Pedestrian detection from a moving vehicle", *Proc. 6th European Conf. Computer Vision*, Vol. 2, pp. 37–49.

[18] Stauffer, C. & Grimson, W.E.L. (2000) "Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 747-757.
[19] Kaewtrakulpong, P. & Bowden, R. (2001) "An improved adaptive background mixture model for real-time tracking with shadow detection", *Proc. 2nd European Workshop Advanced Video-based Surveillance Systems*. pp. 149-158.
[20] Welch, G. & Bishiop, G. (2006) *An Introduction to Kalman Filter*, Department of Computer Science, University of North Carolina at Chapel Hill.

AUTHORS

Chi-Cheng Cheng was born in Taipei, Taiwan, R.O.C. He received the B.S. degree and the M.S. degree in power mechanical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1981 and 1983, respectively, and the Sc.D. in mechanical engineering from Massachusetts Institute of Technology, Massachusetts, USA, in 1991. He is currently a Professor with the Department of Mechanical and Electro-Mechanical Engineering of the National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C. His research interests are in the areas of systemdynamics and control, machine vision, intelligent robots, and mechatronics.

Yi-Fan Wu was born in Qinhuangdao, Hebei, China. He obtained the B.S. degree in mechanical engineering from Southwest Jiaotong University, Chengdu, Sichuan, China in 2014 and the M.S. degree in mechanical and electro-mechanical engineering from National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C. in 2016. His research interests include machine vision and automatic control.