Journal of Information Hiding and Multimedia Signal Processing ©2012 ISSN 2073-4212
Ubiquitous International Volume **3**, Number **1**, January **2012**

A Cost-Effective People-Counter for a Crowd of Moving People Based on Two-Stage Segmentation

Chao-Ho Chen[1], Tsong-Yi Chen[1], Da-Jinn Wang[2] and Tsang-Jie Chen[1]

Department of Electronic Engineering[1]
National Kaohsiung University of Applied Sciences
No.415, Chien Kung Rd., Kaohsiung 807, Taiwan(R. O. C.)
thouho@kuas.edu.tw

Department of Information Management[2]
National Kaohsiung Marine University
No.142, Haijhuan Rd., Nanzih District, Kaohsiung 81143, Taiwan(R. O. C.)
wangdaj@mail.nkmu.edu.tw

Received May 2011; revised November 2011

Abstract. *This paper is dedicated to a cost-effective people counter for a crowd of moving people by using a zenithal video camera. To obtain a more accurate people count, the two-stage segmentation is developed for extracting each person from a crowd. Firstly, a crowd is segmented by frame-difference technique, followed by morphological processing and region growing. Then, the connected-component labeling method is used to generate many individual people-patterns from the segmented crowd. People-image features, such as the area, height, and width of each people-pattern, are analyzed in order to correctly segment each person from each individual people-pattern. Finally, each person segmented is tracked till touching the base-line and then is counted. Experimental results show that the counting accuracy can be achieved above 91% on average if the crowd moves normally. A comparison with other reported methods of using a zenithal camera manifests the superiority of the proposed method in counting accuracy.*
Keywords: People counter, Crowd, Video processing, Moving-object segmentation

---

1. **Introduction.** An accurate automatic people counter is very attractive for the entry control and access surveillance of the important military, building security and commercial applications. Anyway, the early automatic people counting systems, such as turn stiles, rotary bar, and light beams, cant accurately count the people flow when there is more than one person passing through a gate or door at one time. To overcome this problem, many image-processing based approaches with various designs [1]-[14] are hence motivated and they all provide a real-time automatic counting for passing people through a specific region of interest by analyzing a series of images captured with the video camera.

For the transportation applications, Bartolini et al. [1] and Albiol et al. [2] addressed the problems of determining the number of people getting into and out of a bus and train, respectively. To avoid the occlusion problem, Rossi and Bozzoli [3] and Sexton et al. [4]

mounted the camera vertically with respect to the floor plane and set the optical axis of the camera in such a way that the passing people could be observed from just overhead. Though, the system [3] based on template motion-estimation tracking may be very time-consuming because the computation complexity increases substantially with the increasing number of pedestrians and it may suffer from people-touching overlapping problem. Focused on dynamic backgrounds, Zhang and Sexton [5] developed an automatic pedestrian counting method on an escalator or a moving walkway by using a model-specified directional filter to detect object candidate locations followed by a novel matching process to identify the pedestrian head positions in an image even with complicated contents. With the graylevel-based head analysis, the method will suffer from the following situations: a low contrast of the head image with the background and hair styles or wearing various hats for pedestrians. The first case illustrates that the graylevel technique cant provide sufficient information for extracting the required pattern from an image, and the second case reveals that a model-based processing may be affected by various sizes and shapes of the human body due to clothing. To increase the count of passing people through a gate at one time, Terada et al. [6] used the stereo images captured by a pair of cameras to cope with both problems of the crowd counting and direction recognition of the passing people. Anyway, the setting of the stereo camera is complicated and the measurement will be seriously sensitive to any shift of camera. To avoid limiting the setting position of the camera and re-counting someone people as they move around, the approach of using multiple cameras located over the region of interest will be an allowable solution [7, 8]. Based on the cost-effective consideration, a single camera with a tracking algorithm may be the better solution and thus Masoud and Papanikolopoulos [9] developed a rectangular model-based recognition of the pedestrian with human motion analysis to achieve a reliable people count. By setting a fixed single camera hung from the ceiling of the gate, Kim et al. [10] proposed a real-time scheme to detect and track the people moving in various directions with a bounding box enclosing each person. Also using a single zenithal camera, Bescos et al. [11] introduced a DCT based segmentation, which can efficiently consider both lighting and texture information, to cope with some problems, such as shadows, sudden changes in background illumination and sporadic camera motion due to vibration, in order to count people crossing an entrance to a big store. Based on top-view video sequences, a people counting system can be used in different illumination conditions by employing region merging to remove shadows of each object that is segmented by k-means clustering [12]. For counting the passengers getting in/out of a bus, the captured frame by a zenithal camera is firstly divided into many blocks and each block will be classified according to its motion vector. If the accumulated number of blocks with similar motion vectors is greater than a threshold, those blocks are regarded as belonging to the identical passenger pattern. Such a pattern is then judged if it is a passenger to be counted [13]. For solving the frequently-happened overlapping problem [14], called people-image overlapping which is mainly resulted from people touching with each other, both area and color information of people are utilized to count the people flow passing through a gate or door. It also copes with the merge-split problem that people walk sometimes touching with one another and sometimes separating from others. However, the above people counting methods havent presented a solution or cant provide an accurate count for a crowd of moving people.

In another approach, by taking advantage of human motion analysis, many techniques of the human body tracking or pedestrian detection [15-26] may be applied to the pedestrian counting in open spaces, in which the camera is usually set with a downward-slope view to obscure a more sufficient surveillance range. Nevertheless, the tracking process is always very computation-intensive and their camera setting will make it is difficult to segment or recognize each person in the crowded pedestrians owing to a serious overlapping problem.

For the purpose of improving the counting accuracy for a crowd of moving people in various illumination conditions, a two-stage-segmentation based people-counting method using a zenithal video camera is proposed. In the proposed method, a crowd is first segmented by frame-difference technique, followed by morphological processing and region growing. Then, the connected-component labeling method is used to generate many individual people-patterns from the segmented crowd. People-image features, such as the area, height, and width of each people-pattern, are employed in order to correctly segment each person from each individual people-pattern. Finally, each person segmented is tracked till touching the base-line and then is counted. The following section will describe the proposed people-counting algorithm. Then, experimental results, analysis, and comparison are discussed in Section 3 and conclusions are made in the final section.

2. **The Proposed People-Counting Algorithm.** In the proposed system setting, the camera is set with a downward viewing and hence it has the least affection of people-image overlapping. Figure 1 describes the proposed people counting algorithm which mainly includes crowd segmentation, person segmentation, and person counting and tracking.

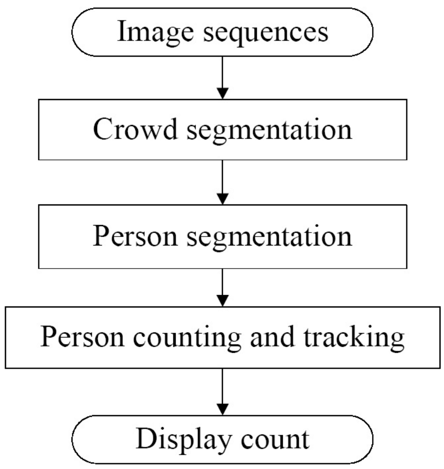

FIGURE 1. The proposed people counting algorithm.

2.1. **Crowd Segmentation.** Basically, the background subtraction technique is not suitable for moving-object segmentation if the illumination condition is changeable. The major reason is that the background cant be removed completely. Though the optical-flow approach may overcome the above problem, it suffers a computation-intensive problem and thus will not suitable for the real-time applications. The frame-difference approach can also avoid the above problem, but the extracted moving-object region only appears a rough shape of that object, not the whole body. In the proposed crowd segmentation, the frame-difference technique is first employed to segment the moving-object and then the morphological processing is utilized to obtain a complete shape of each significant moving-object. A region-growing technique is further used to fill the various holes within each moving-object. To refine each body extracted in the above, the color-based body compensation is exploited to derive an optimal moving-object mask. The proposed crowd segmentation process is shown in Figure 2.

As mentioned above, frame-difference will generate a rough hollow body of each pedestrian. To obtain a complete body, dilation-based morphological processing is first used to enhance the boundary and then region-growing is used to fill the holes within each body as possible. Figure 3 describes the process of the region-growing algorithm, where red block denotes a selected seed in the subfigure (b), eight-connection growth (red blocks) is adopted as shown in the subfigure (c), the final result (red blocks) of region-growing is depicted in the subfigure (d). Anyway, from the above region-growing result, there are still some holes existed in each body. For generating a complete body, a refinement process using color-based compensation is developed in order to further fill those residual holes after executing the region-growing algorithm.

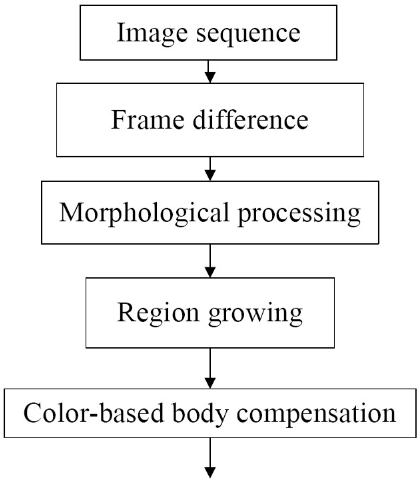

FIGURE 2. The proposed crowd segmentation process.

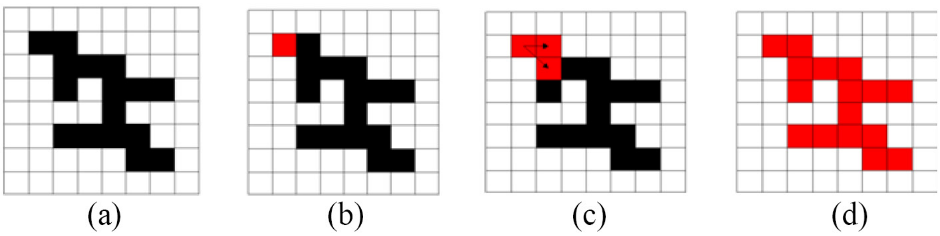

FIGURE 3. The process of the region-growing algorithm: (a) Origin region; (b) Seed selection; (c) Eight-connection growth; (d) Result of region-growing.

For the frames of region-growing result, let $D_n(x,y)$ denotes the union of differences between the current frame $RF_n(x,y)$ and the previous k frames $RF_{n-j}(x,y)$, $j = 1, 2, ...k$, as illustrated in equation (1). Based on HSV (Hue-Saturation-Value) color model, $DH_n(x,y)$ is defined as hue values of pixels in $D_n(x,y)$, $BH(x,y)$ is defined as hue values of background pixels at the same pixel position (x,y) for those previous k frames, and $DDH_n(x,y)$ is denoted as hue difference between $DH_n(x,y)$ and $BH(x,y)$, as shown in equation (2). If the hue difference $DDH_n(x,y)$ is larger than a color tolerance (CT) value, the pixel of position (x,y) is regarded as foreground pixel (i.e., moving-object pixel). These foreground pixels are then utilized to compensate for those residual holes within the moving-object, as described in equation (3), in which the union of $RF_n(x,y)$ and $OF_n(x,y)$ is the final compensated result $CF_n(x,y)$, where $OF_n(x,y)$ denotes the pixel of original frame.

$$D_n(x,y) = \bigcup_{j=1}^{k} |RF_n(x,y) - RF_{n-j}(x,y)| \tag{1}$$

$$DDH_n(x,y) = |DH_n(x,y) - BH(x,y)| \tag{2}$$

$$CF_n(x,y) = \begin{cases} RF_n(x,y) \cup OF_n(x,y), & DDH_n(x,y) > CT \\ RF_n(x,y), & DDH_n(x,y) \le CT \end{cases} \tag{3}$$

Figure 4 demonstrates an example of crowd segmentation through performing the above processes, where subfigure (a) is the original frame, subfigure (b) is the result of frame-difference, subfigure (c) is the result of morphological processing, subfigure (d) is the result of region-growing, subfigure (e) is the compensation region, and subfigure (f) is the final result of crowd segmentation. In the subfigure (b), the frame-difference processing generates an incomplete body-image with many various holes, though the background and shadow are removed substantially. The morphological processing effectively reduces the deficiencies of bodies resulted from the frame-difference processing, but there are some various holes existed in bodies, as shown in the subfigure (c). From the subfigure (d), the region-growing cant completely remove all holes within each body due to various gaps which will break the growths of those seeds. By using the color-based compensation method, the subfigure (e) shows the compensation region (i.e., white area) that is dedicated to fill those residual holes. Finally, the compensated body-images are described in the subfigure (f).

2.2. **Person Segmentation.** When a moving crowd is segmented, it is necessary to identify each body for the following counting purpose. From the segmented crowd, the connected-component labeling method is first used to generate many individual objects and then people-image features are analyzed for judging how many persons each individual object contains. If the extracted object contains at least two persons, each person needs to be segmented within this object.

Basically, the connected-component labeling [13] is used to collect the connected pixels with similar features to be an individual object. In the proposed method, the segmented crowd image is first transformed to a binary image and then a 3*3 mask is introduced to label the neighboring pixels with similar value, where different groups of connected pixels will be labeled respectively. At last, the pixel with the same label will form an individual object and thus many individual objects are yielded. If the area (i.e., pixel number) of an individual object is smaller than a threshold TSP, i.e., typical area of a single person, this object is considered to be not people-pattern and thereby it should be removed from the segmented crowd mentioned above. Figure 5 describes the proposed connected-component labeling algorithm for labeling all people-patterns.

After labeling each people-pattern in a segmented crowd, an initial count can be obtained, as shown in Figure 6, where each people-pattern is bounded with a rectangle. In the figure, the labeling order is the row-wise order, i.e., scanning from right to left and top to bottom. The lack of label-number 3, 5, 6, 8-17 implies these objects (i.e., group of connected pixels) have been removed because each objects area is smaller than TSP. In Figure 6, there are only 5 objects left, anyway, each object will contains at least one person and thus it is necessary to determine the number of persons existed in each object for the final counting.

To analyze the people-pattern number of an object, pixel quantity, height, and width of the object are utilized. At first, each object is bounded with a rectangle (as shown in Figure 6) and then both width and height of such a bounded-box and the objects area (i.e., pixel number) are exploited to deduce the number of people-patterns within this object. In the deduction, SP denotes the standard pixel number of a person, P_{Pixels} denotes the pixel number of an object, P_{Width} and P_{Height} denote the width and height of a bounded-box, respectively. If $P_{Pixels} > TSP$, it reveals that this object will contain more than one people-pattern. Hence, this object will be divided into several subobjects on average and then each subobject will be also checked if it contains more than one people-pattern. If the pixel number of a subobject is larger than TSP, this subobject will be averagely divided into several subobjects again. In the dividing process, each split is an

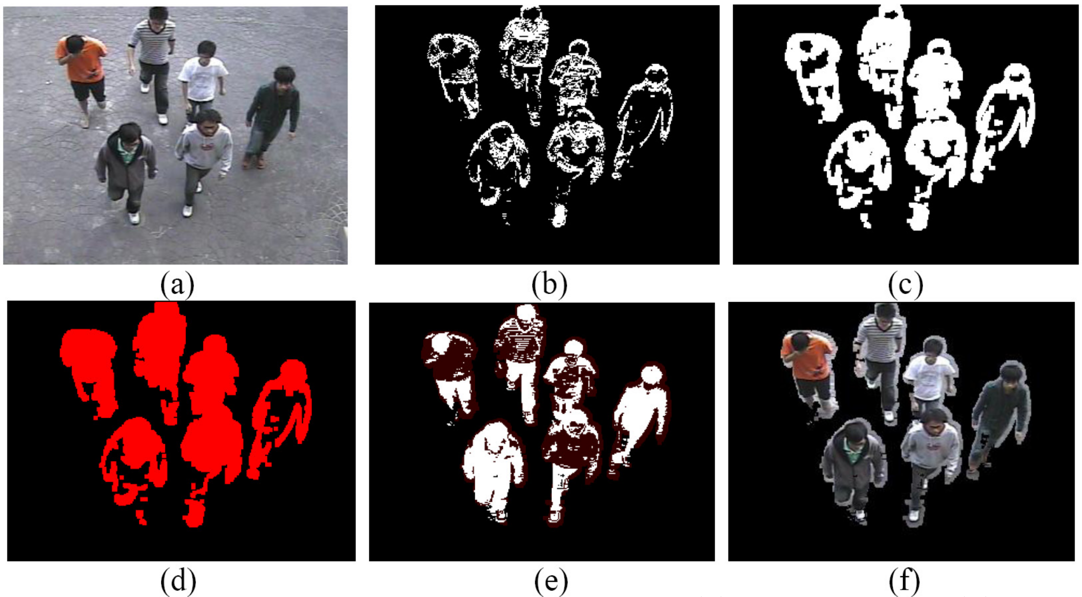

FIGURE 4. An example of crowd segmentation: (a) Original frame; (b) Result of frame-difference; (c) Result of morphological processing; (d) Result of region-growing; (e) Compensation region; (f) Final crowd segmentation result.

average division and this division process is performed iteratively, till the pixel number of a divided-object is smaller than TSP. During the iterative dividing process, the division is realized by splitting the bounded-box in the order of from width to height of a bounded-box. From the above analysis, the quantity of people-patterns within a bounded-box is deduced by using the following rule, where TW and TH denote the threshold values of width and height of a people-pattern, respectively, and Round() is a rounding function.

Table 1 lists the values of P_{Pixels}, P_{Width}, and P_{Height} of every people-pattern in Figure 6, where the captured video format provides a resolution of 320*240 at 30 frames/sec. Basically, the shape of people-pattern is variable because human motion is not rigid. In practice, those thresholds of TSP, TW and TH should be set according to real situation of applications. With $TSP = 4200$, $TW = 75$ and $TH = 120$, Figure 7 shows the division result of using the above dividing rule for Figure 6. In the figure, only people-pattern of label-4 is bisected into both new people-patterns of label-4 and label-29, due to the fact that the pixel number (6133) and P_{Height} (178) of people-pattern of label-4 are larger than $TSP(4200)$ and $TH(120)$, respectively, from Table 1. The new values of P_{Pixels}, P_{Width}, and P_{Height} of every people-pattern in Figure 7 are listed in Table 2, where P_{Pixels}, P_{Width}, and P_{Height} of people-pattern of label-4 are reduced to 3290, 54, and 89, respectively, and those values of the added people-pattern of label-29 are 2792, 58, and 88. From Table 2, there is no any value of P_{Pixels}, P_{Width}, and P_{Height} for every people-pattern to be larger than those thresholds of $TSP = 4200$, $TW = 75$ and $TH = 120$, respectively. As a result, the number of people-patterns with label-1, 2, 4, 7, 18, and 29 implies people $count (= 6)$ at that time.

2.3. **Person Counting and Tracking.** From the above person segmentation, each people-pattern has been labeled for the following counting. In the moving process for a crowd of people, those people-patterns may generate the merge-split problem [13]. To determine if each person will be counted, the people-pattern requires to be tracked till it touches a base-line. In general, the crowd always moves so slowly and hence the bounding-boxes intersection-checking technique can be employed for tracking the people-pattern. A

If $P_{Pixels} > TSP$, then

If $P_{Width} > TW$, then

n = Round(P_{Width}/TW), Divide object into n parts on average

If $P_{Height} > TH$, then

n = Round (P_{Height}/TH), Divide object into n parts on average

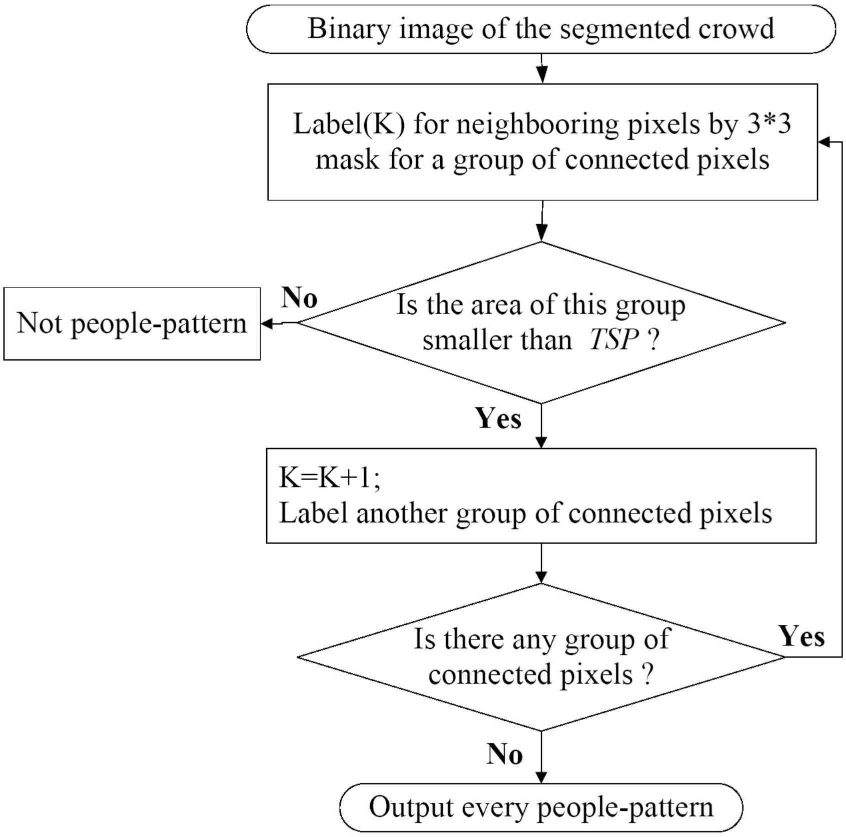

FIGURE 5. The proposed connected-component labeling algorithm.

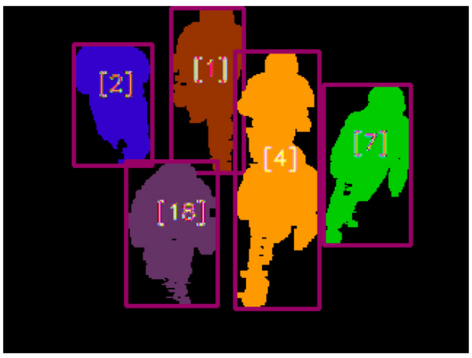

FIGURE 6. Result of the proposed connected-component labeling for Figure 4 (f).

bounding-box is introduced to trail the people-pattern, as illustrated in Figure 8, where an intersectional case is depicted. It should be noted that the value of time interval n can be

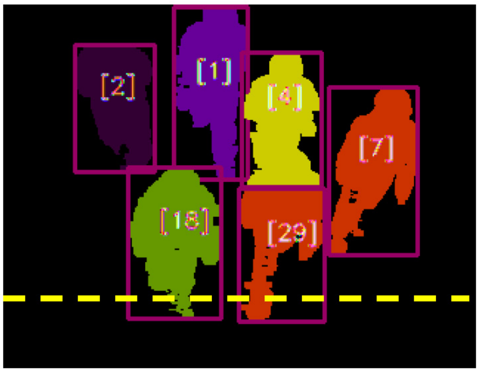

FIGURE 7. The division result of using the proposed dividing rule for Figure 6.

TABLE 1. The values of P_{Pixels}, P_{Width}, and P_{Height} of every people-pattern in Figure 6.

Label	P_{Pixels}	P_{Width}	P_{Height}
1	3515	50	114
2	3017	54	84
4	6133	58	178
7	3584	60	111
18	3778	63	100

TABLE 2. The values of P_{Pixels}, P_{Width}, and P_{Height} of every people-pattern in Figure 7.

Label	P_{Pixels}	P_{Width}	P_{Height}
1	3514	50	114
2	3019	54	84
4	3290	54	89
7	3584	60	111
18	3780	63	100
29	2792	58	88

adjustable and dependent on the frame-sampling rate. If there is an intersection existed in two bounding-boxes of the identical people-pattern in consecutive frames, it is judged that such two people-patterns in these boxes belong to the same person. On the contrary, i.e., their bounding-boxes have no intersection, it may generate an abrupt situation and thus the two separated people-patterns will not represent the identical person. So, this tracking will be aborted and another new tracking is started. The above description of the tracking procedure is illustrated in Figure 9.

In respect of tracking, the intersection check of the bounding-box pair for person-identification will last throughout the whole tracking process till the person is counted, i.e., touching the base-line (yellow line), as shown in Figure 7. In the normal situation, the above tracking method can cost-effectively catch each person in a crowd. Anyway, there may be some special situations which will make the two bounding-boxes to be disjointed, such as someone people moves fast, turns back, or even makes an intentional disturbance. These cases will make the tracking procedure to be aborted and thus a new people-pattern tracking is started.

3. **Experimental Results and Discussions.** A theoretical analysis about the proposed counter for a crowd of moving people has been given in the above section, but the implementation of such a counting system with the practical image sequences can provide a realistic and interesting evaluation. To realize the proposed people-counting system,

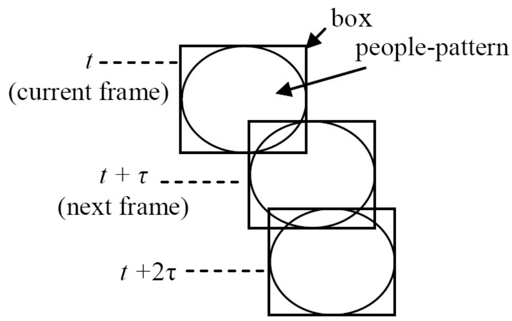

FIGURE 8. Trailing of a bounding-box for intersectional case.

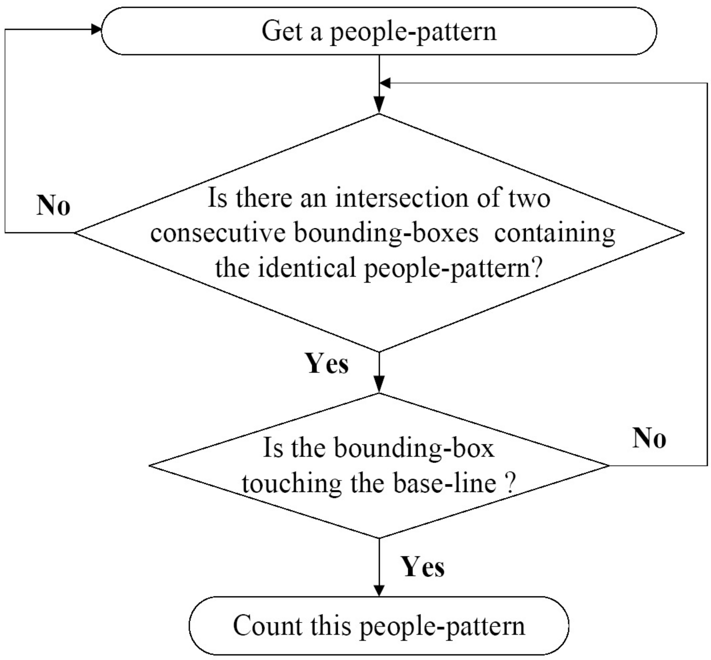

FIGURE 9. The tracking procedure.

a video camera (TS-730H) is set 3.6m above the floor, with a surveillance area of 320 240 pixels (i.e., frame size). The capture rate of the camera is 30 (frames/second) but the frames processed per second ranging from 20 to 30, depending on the people number of a crowd. For the deep evaluation, six image sequences, Video-A-1, Video-A-2, Video-A-3, Video-B-1, Video-B-2, and Video-B-3, which are captured at two different places of site-A (outdoors) and site-B (indoors), are used for the input to the proposed counting algorithm. It is noted that the major difference of such two places is their different shadow effects, where the shadow effect of site-A is smaller than that of site-B. In these sequences, the crowds moving directions include uni-direction and bi-direction under a moderate moving speed (not more than 80 meters per minute), where both crowds directions in Video-A-1 and Video-A-2 are uni-directional but opposite, the crowds direction in Video-A-3 is bi-directional, and the crowds directions in Video-B-1, 2, 3 are the same as those of Video-A-1, 2, 3, respectively. Figure 10 shows their 680-th frames of these six sequences.

To manifest the superiority of the proposed people-counting method to four reported methods of [11], [12], [13], and [14], a comparison in terms of count accuracy through using the above six sequences, is made in Table 3. In a fair comparison, those reported methods should be realized by a zenithal-camera setting scheme as same as the proposed method. From the Table, it is clear that the proposed method provides a higher count accuracy of 91% on average, due to the fact that the people-pattern overlapping caused by crowding can be solved substantially. For these compared methods, Video-B-1 generates the best accuracy than other test sequences because this sequence contains less crowding people. Relatively, Video-B-3 generates the smallest accuracy than other test sequences because this sequence contains crowded and bi-directional moving people, which will make person segmentation to be more difficult. On average, the proposed method provides the counting accuracy (91.67%) that exceeds those of other three previous methods [11], [12] and [13], over 20%, and about 10% over that of [14]. This is because these three previous methods havent solved the crowding problem. Though the counting method of [14] has proposed area-based people number estimation for a crowd of people and thus it achieves accuracy of 82.83%, it may fail if both crowding and shadow come up in the meantime.

From an overall analysis on the above experimental results, it implies that it is difficult to segment a people-pattern for the bi-directional crowded moving people and this will become more intractable when the above case involves shadow influence. This is major reason of confusing the count in the proposed method. In addition, it is also arduous to segment each person from a big people-pattern which is resulted from very crowding situation. However, the proposed people counting method is still very effectively for a crowd of moving people excluding the abrupt change of peoples movement and intentional actions.

4. **Conclusions.** This paper presents an automatic bi-directional people-counting method for a crowd of moving people by segmenting each person from a crowd. To improve the inherent rough estimation of people number within a big people-pattern caused by the crowding situation, two-stage segmentation is developed for providing a more reliable people count in a crowded people. Experimental results show that the counting accuracy can be achieved above 90% if the crowd moves normally. When compared to other methods of using a single zenithal camera setting scheme, the proposed people counting method can obtain higher counting accuracy for a crowd of moving people. Therefore, the proposed method will provide a cost-effective people-counting technique and thus it will be more attractive than other methods for counting crowded moving people.

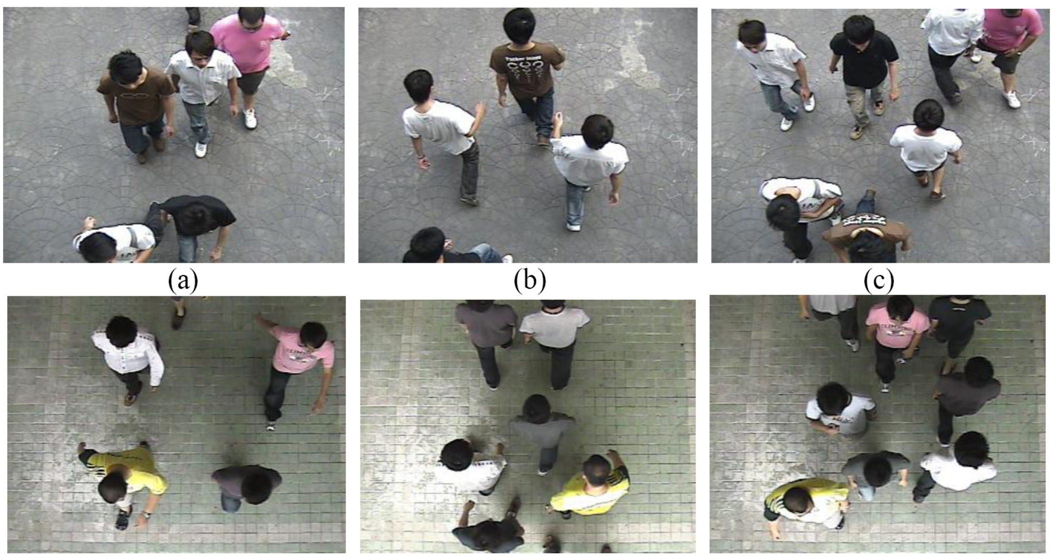
(a) (b) (c)
(d) (e) (f)

Figure 10. The 680-th frames of these six sequences: (a) Video-A-1; (b) Video-A-2; (c) Video-A-3; (d) Video-B-1; (e) Video-B-2; (f) Video-B-3.

Table 3. Comparison of count accuracy for five methods using six sequences.

Sequence (400 sec)	Real people number		Count		[11]	[12]	[13]	[14]	The proposed
	In	Out	In	Out	Accuracy				
Video-A-1	0	119	0	105	66%	66%	70%	85%	94%
Video-A-2	78	0	83	0	67%	68%	69%	82%	93%
Video-A-3	70	74	84	75	68%	71%	72%	83%	91%
Video-B-1	0	103	0	105	83%	85%	87%	85%	95%
Video-B-2	104	0	98	0	65%	67%	67%	82%	91%
Video-B-3	46	45	57	48	62%	64%	63%	80%	87%
Average					68.5%	70.17%	71.33%	82.83%	91.67%

Acknowledgment. This work is supported by National Science Council under Grant NSC 99-2622-E-151-015-CC3, Taiwan, R.O.C.

REFERENCES

[1] F. Bartolini, V. Cappellini and A. Mecocci, Counting people getting in and out of a bus by real time image-sequence processing, *Image and Vision Computing*, vol. 12, no. 1, pp. 36-41, 1994.

[2] A. Albiol, I. Mora and V. Naranjo, Real-time high density people counter using morphological tools, *IEEE Trans. Intelligent Transportation Systems*, vol. 2, no. 4, pp. 204-218, 2001.

[3] M. Rossi and A. Bozzoli, Tracking and counting moving people, *Proc. of IEEE International Conference on Image Processing*, vol. 3, pp. 212-216, 1994.

[4] G. Sexton, X. Zhang and G. Redpath, Advances in automatic counting of pedestrians, *Proc. of the European Convention on Security and Detection*, pp. 106-110, 1995.

[5] X. Zhang and G. Sexton, Automatic human head location for pedestrian counting, *Proc. of the IEE 6th International Conference on Image Processing and its Applications*, pp. 535-540, 1997.

[6] K. Terada, D. Yoshida, S. Oe and J. Yamaguchi, A method of counting the passing people by using the stereo images, *Proc. of IEEE International Conference on Image Processing*, vol. 2, pp. 338-342, 1999.

[7] L. Khoudour, J. P. Deparis and L. Durvieubourg, Linear image sequence analysis for passengers counting in public transport, *Proc. of IEE International Conference on Public Transport Electronic Systems*, pp. 100-104, 1996.

[8] V. Kettnaker and R. Zabih, Counting people from multiple cameras, *Proc. of IEEE International Conference on Multimedia Computing and Systems*, vol. 2, pp. 267-271, 1999.

[9] O. Masoud and N. P. Papanikolopoulos, A novel method for tracking and counting pedestrians in real-time using a single camera, *IEEE Trans. Vehicular Technology*, vol. 50, no. 5, pp. 1267-1278, 2001.

[10] J. W. Kim, K. S. Choi, B. D. Choi and S. J. Ko, Real-time vision-based people counting system for security door, *Proc. of International Technical Conference on Circuits/Systems Computers and Communications*, pp. 1416-1419, 2002.

[11] J. Bescos, J. M. Menendez and N. Garcia, DCT based segmentation applied to a scalable zenithal people counter, *Proc. of IEEE International Conference on Image Processing*, vol. 3, pp. 1005-1008, 2003.

[12] J. M. Wang, L. K. Lee, Y. C. Chung and S. W. Chen, People counting based on top-view video sequence, *Proc. of the 18th IPPR Conference on Computer Vision, Graphics and Image Processing*, pp. 200-207, 2005.

[13] C. H. Chen, Y. C. Chang, T. Y. Chen and D. J. Wang, People counting system for getting in/out of a bus based on video processing, *Proc. of the 8th IEEE International Conference on Intelligent Systems Design and Applications*, pp. 565-569, 2008.

[14] T. Y. Chen, T. H. (C. H.) Chen and D. J. Wang, A cost-effective people-counter for passing through a gate based on image processing, *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 3, pp. 785-800, 2009.

[15] J. K. Aggarwal and Q. Cai, Human motion analysis: A review, *Proc. of IEEE Conference on Nonrigid and Articulated Motion*, pp. 90-102, 1997.

[16] B. Heisele and C. Wohler, Motion-based recognition of pedestrians, *Proc. of the 14th International Conference on Pattern Recognition*, vol. 2, pp. 1325-1330, 1998.

[17] I. Haritaoglu, D. Harwood and L. S. Davis, Real-time surveillance of people and their activities, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809-830, 2000.

[18] O. Masoud and N. P. Papanikolopoulos, Robust pedestrian tracking using a model-based approach, *Proc. of IEEE International Conference on Intelligent Transportation Systems*, pp. 338-343, 1997.

[19] T. Darrell, G. Gordon, M. Harville and J. Woodfill, Integrated person tracking using stereo, color and pattern detection, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 601-608, 1998.

[20] C. Bregler, Learning and recognizing human dynamics in video sequences, *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 8-15, 1998.

[21] S. Wachter and H. H. Nagel, Tracking persons in monocular image sequences, *Proc. of Computer Vision and Image Understanding*, vol. 74, no. 3, pp. 174-192, 1999.

[22] Masakatsu Mitani, Mamoru Takaya, Atsuhiro Kojima and Kunio Fukunaga, Dynamic recognition of human actions and objects using dual hierarchical models, *International Journal of Innovative Computing, Information and Control*, vol. 3, no. 6, pp. 1359-1368, 2007.

[23] T. Y. Chen, T. H. (C. H.) Chen, D. J. Wang and Y. C. Chiou, Real-time video object segmentation algorithm based on change detection and background updating, *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 7, pp. 1797-1810, 2009.

[24] D. J Wang, C. H. Chen and C. T. Lee, People recognition for entering and leaving a video surveillance area, *Journal of Software*, vol. 5, no. 12, pp. 1342-1348, 2010.

[25] C. H. Chen, T. Y. Chen, D. J. Wang and Y. F. Li, Multipath flatted-hexagon search for block motion estimation, *Journal of Information Hiding and Multimedia Signal Processing* , vol. 1, no. 2, pp. 110-131, 2010.

[26] W. C. Hu, C. Y. Yang, D. Y. H. and C. H. Huang, Feature-based face detection against skin-color like backgrounds with varying illumination, *Journal of Information Hiding and Multimedia Signal Processing* , vol. 2, no. 2, pp. 123-132, 2011.