ICIC Express Letters
Part B: Applications
Volume **7**, Number **12**, December **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **2563–2568**

REAL-TIME PEOPLE COUNTING SYSTEM BASED ON STEREO VISION

Hsiang-Yueh Lai* and Meng-Hua Sung

Department of Electrical Engineering
National Chin-Yi University of Technology
No. 57, Sec. 2, Chung-Shan Rd., Taiping Region, Taichung City 41101, Taiwan
*Corresponding author: anne@ncut.edu.tw

Received May 2016; accepted August 2016

Abstract. *People counting systems are usually positioned in the doorway or a corridor of a building. Turnstiles, optical sensors and image processing are often used as people counting system. In this paper, a real-time people counting system based on stereo vision is proposed. Two top view cameras are used to capture images. Stereo vision algorithm and sum of the absolute difference algorithm are adopted to detect people by human characteristics. The human height and the shoulders are found to detect people correctly. The people counting system effectively reduces people image occlusion and shadow problems, and successfully solves the miscount problems of vision count systems which mistake a goods as a person. The system counts people number with high accuracy.*
Keywords: Stereo vision, People counting, People tracking

1. **Introduction.** A people counting system has often been used in a video monitor system. The average daily customer flow is an important message for a public transit or a service industry company. That supports managers to make corresponding adjustments of strategies. A people counting system is usually positioned in the doorway or a corridor of a building.

The people counting systems with turnstiles reduce the people flow speed when people go through turnstiles [1]. The people counting systems with optical sensors have view occlusion problems when people go through optical sensors side by side. The optical sensor systems may consider an object as a person and have error count information. Traditional surveillance systems with manual monitoring have the best accuracy. However, when a person monitors multiple screens for a long time, he will be tired and cause a miscount. A computer vision system is more stable than a manual monitoring system. The counting system based on human face detection also has view occlusion problems [2]. The camera with a bevel scene can catch more human characteristics but has view occlusions for people counting [3]. The camera positioned on overhead can solve occlusion problems but has lower human characteristics [4-7].

We use two top view cameras to capture images. The targets are detected by using the background subtraction method and the stereo vision algorithm. That solves the image occlusion and shadow problems. The stereo vision and the sum of absolute difference (SAD) algorithm calculate people's heights and shoulders as human characteristics to perform the real-time people counting system based on stereo vision.

2. **System Architecture.** The people counting system framework is shown in Figure 1. The hardware of our people counting system is shown in Figure 1(a). Two top view cameras capture analog images, and then analog image signals are converted to digital image signals by a video capture card. The digital image signals are processed by a computer. Figure 1(b) is the system flow chart of our system. First, motion targets are

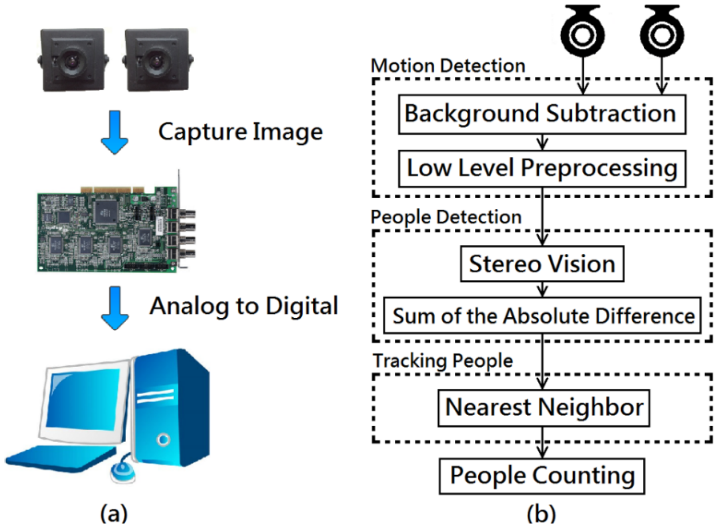


FIGURE 1. People counting system framework

found by background subtracting and low level preprocessing. Then people in motion are detected by the stereo vision and the SAD algorithm. The nearest neighbor (NN) algorithm is used to track people in motion and count the number of people.

3. Motion Detection.

3.1. Background subtraction. The background removal method detects all different backgrounds and moving targets, especially when targets move in a stop and go pattern. The background removal method is shown in Equation (1).

$$I_{motion}(m,n) = \begin{cases} I_{src}(m,n) & \text{if } (I_{src}(m,n) - I_{BG}(m,n)) \geq Thr \\ 0 & \text{else} \end{cases} \tag{1}$$

where I_{motion} is a motion detection image, I_{src} is a camera capturing image, I_{BG} is a background image, and Thr is the threshold value. When $I_{src}(m,n) - I_{BG}(m,n)$ is greater than or equal to Thr, $I_{motion}(m,n)$ is equal to $I_{src}(m,n)$; else output pixels are black. Our system uses color images in order to increase the accuracy of SAD algorithm.

3.2. Low level preprocessing. After removing the background, the foreground image is converted to a grayscale image, and the output is a threshold image as Equation (2).

$$I_{pre}(m,n) = \begin{cases} 255 & \text{if } I_{Gray}(m,n) > 0 \\ 0 & \text{else} \end{cases} \tag{2}$$

where I_{pre} is the threshold image, and I_{Gray} is the grayscale image. In order to get a clear image, we adopt the morphological opening to remove small noise, and then morphological closing to fill the voids.

4. People Detection and Counting.

4.1. Stereo vision. A three-dimensional image is formed by two different visual images as the characters of human eyes. It means that the information of a target can be calculated by two images with different visual angles captured by two cameras. Figure 2 shows a schematic diagram of the epipolar geometry of stereo vision.

In Figure 2, P is a target, O_l is the left camera center, O_r is the right camera center, the dash line through x_l and x_r is the horizon of two cameras imaging, x_l and x_r are the target X-axis coordinates of the left and right cameras imaging on the dash line $\overline{x_l x_r}$ plane, B is the distance between the two camera centers, and f is the focal length of

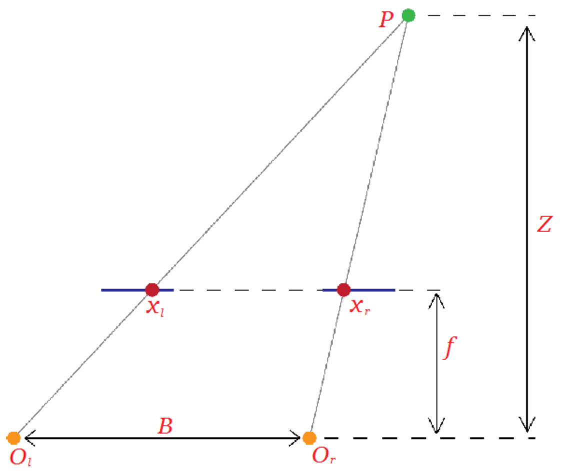


FIGURE 2. Schematic diagram of epipolar geometry of stereo vision

the two cameras. The two cameras are set up in the same horizon, i.e., the same Y-axis, so it is only necessary to consider the X-axis and Z-axis. The parallax is defined as $d = x_l - x_r$. Parallax d can be calculated by image processing techniques with two visual angle images. Equation (3) is calculated by the similar triangles in Figure 2. Then Equation (4) calculates the height Z of target P.

$$\frac{B-(x_l-x_r)}{Z-f} = \frac{B}{Z} \tag{3}$$

$$Z = \frac{fB}{d} \tag{4}$$

4.2. **Sum of the absolute difference.** A stereo vision image has the property that an object appears in one side of the visual angle view image and usually appears in the other image. First, we cut a visual angle view image into many blocks with SAD in Equation (5), where the column number and row number are m and n of a search block. And then, we find the block in right visual angle view image which has the most similar block in left visual angle view image. The displacement of the blocks with minimum difference is recorded for stereo vision.

$$V_{SAD} = \sum_{j=0}^{m=1}\sum_{i=0}^{n=1} abs[I_1(i,j) - I_2(i,j)] \tag{5}$$

4.3. **People detection.** Figure 3(a) shows pedestrians in and out the counting area. If a turnstile is used in the counting area, the pedestrian pass speed will be reduced.

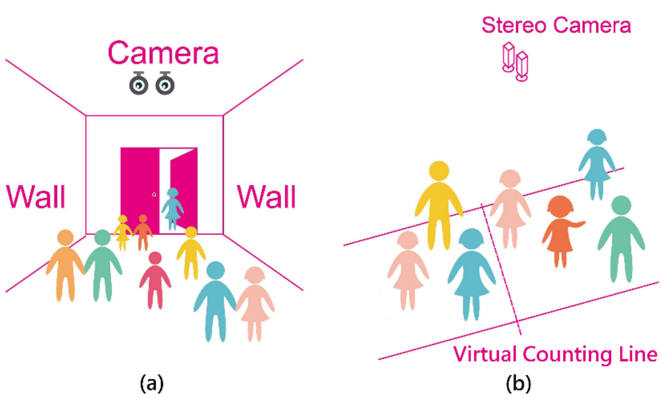


FIGURE 3. (a) Pedestrians through the counting area, (b) schematic diagram of our people counting system

Figure 3(b) shows a schematic diagram of our people counting system. Two top view cameras are set on the ceiling, and capture the top view images when people pass through the counting area. If many people pass through the counting area, the top view images will reduce the overlap of people images.

Images of the two top view cameras are very similar. Figure 4(a) is a source image which shows that the motion targets are moving on the counting area. The counting area is inside the rectangle surrounding the floor. After motion detection processing, motion targets are detected by several virtual measurement lines, as shown in Figure 4(b). The solid line areas are motion targets. When a target is detected by a series of measuring lines, the central pixels of the measure lines are used to determine the target height by stereo vision and SAD. If the target height is greater than a threshold, it may be a person or an object. With the human characteristics, shoulders are found to recognize a person as shown in Figure 4(c). The left target with human shoulders is a person; on the other hand, the right target is not. The highest point of left image is the human height point and the other white points are the human shoulder area around the highest point. The shoulders are found by stereo vision and SAD.

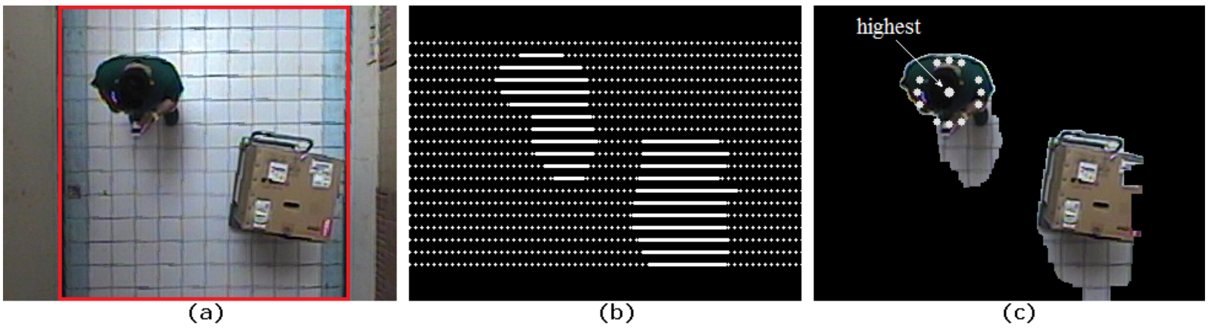


FIGURE 4. (a) A source image, (b) motion targets detection, (c) the human height and shoulders detection

4.4. **Tracking and counting people.** The person in Figure 4(c) is tracked by NN algorithm. For tracking targets, X-axis coordinates and Y-axis coordinates of targets are set as the target characteristics. Equation (6) is Euclidean distance, where Δx and Δy are differences between two frames. The direction of a motion target can be detected by the target characteristics and distance.

$$distance = \sqrt{(\Delta x)^2 + (\Delta y)^2} \tag{6}$$

When a motion person passes inside the visual counting line, the counter number is increased. On the outside direction, the counter number is decreased. When a target leaves the counting area, the system will delete the target and track a new target.

5. **Experimental Results.** Two top view cameras with a resolution 320×240 are set on the ceiling of a build corridor for our system. The distance between the two cameras is 12 cm. The height from ground to the top view cameras is 4.6 meters. The corridor width is 2.435 meters. Our system is carried out on a Windows system.

In the experiments people and objects randomly move through the counting area as shown from Figure 5 to Figure 7. The counter numbers are shown on the upper left corner of the images. The lines are the virtual counting lines. The counting number is increased when people pass through the counting area on the inside, as shown in Figure 5. The counting number is decreased when people pass through the counting area on the outside, as shown in Figure 6. The people counting system successfully solves the image occlusion and shadow problems when pedestrians side by side pass through.

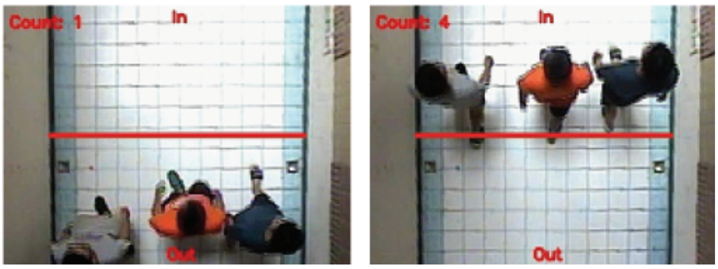


FIGURE 5. People pass through the counting area on the inside.

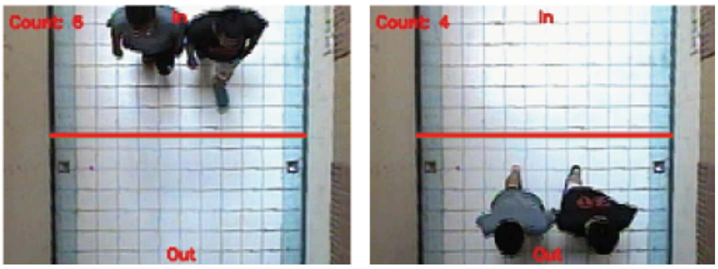
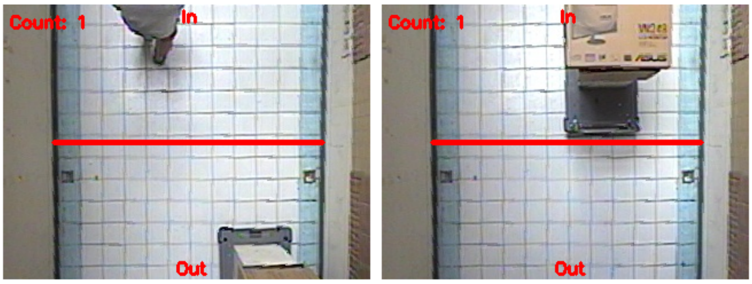
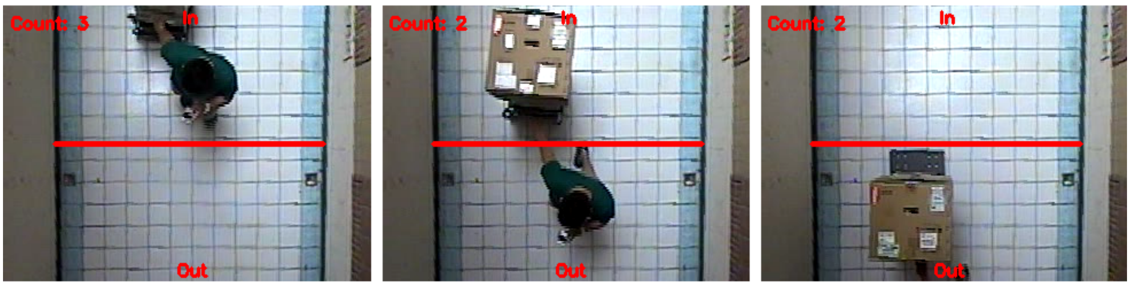


FIGURE 6. People pass through the counting area on the outside.



(a)



(b)

FIGURE 7. (a) Object passes through the counting area; (b) person and object pass through the counting area.

Figure 7(a) shows a goods box on a trolley, totally 1.75 meters in height, passes through the counting area on the inside, the counting system recognizes the motion object as a goods and the counting number does not increase. Figure 7(b) shows that a person with the height of 1.74 meters is dragging a goods with the height of 1.75 meters through the counting area on the outside, and the counting system counts people number correctly. Our system successfully solves the miscount problems of vision counting systems that mistake a goods as a person, and even the goods is nearly of the same height as the person shown in Figure 7(b). Our system can effectively recognize the human characteristics of shoulders and has a good robustness in identification rate.

The identification rates of our system are shown in Table 1. Targets (people or objects) randomly pass through the counting areas. The counting area is the rectangle in Figure

Table 1. Identification rates

Target numbers at the counting area	Identification rate
1	99%
2	98%
3	94%
4 or 4+	84%

4(a). Each experiment condition has been tested 50 times. When target numbers at the counting area are less than 3, the accuracy identification rates are above 98%. When target numbers are equal to or greater than 4 on the counting area, the identification rate is decreased. When targets of more than 4 pass through the small counting area at the same time, the overlaps of the motion targets in the source images reduce the identification rate.

6. **Conclusion.** Computer vision people counting systems usually have the image occlusion and shadow problems, when pedestrians side by side pass through the systems, and the miscount problems of mistaking an object as a person. Based on stereo vision, we propose a real time people counting system. That successfully solves the above mentioned problems. The targets are captured with two top view cameras, and detected by using the background subtraction method and the stereo vision algorithm. That solves the image occlusion and shadow problems and improves the system processing speed. The stereo vision algorithm and SAD calculate the target height. The human height and the shoulders are recognized as human characteristics to identify a person from the detected targets. The motion people are tracked by NN algorithm and the system counts the number of people. The system has high identification accuracy, even when the targets stop and go on the counting area. The system can be further applied in publishing systems, such as mass rapid transit systems, and department stores. That supports managers to make corresponding adjustments of strategies.

REFERENCES

[1] S. Velipasalar, Y. L. Tian and A. Hampapur, Automatic counting of interacting people by using a single uncalibrated camera, *2006 IEEE Int. Conf. Multimedia & Expo*, vol.2006, pp.1265-1268, 2006.

[2] X. Z. X. Zhao, E. Delleandrea and L. C. L. Chen, A people counting system based on face detection and tracking in a video, *The 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, pp.67-72, 2009.

[3] A. S. Rao, J. Gubbi, S. Marusic, P. Stanley and M. Palaniswami, Crowd density estimation based on optical flow and hierarchical clustering, *2013 International Conference on Advances in Computing, Communications and Informatics*, pp.494-499, 2013.

[4] L. Rizzon, N. Massari, M. Gottardi and L. Gasparini, A low-power people counting system based on a vision sensor working on contrast, *Proc. of IEEE Int. Symp. Circuits Syst.*, no.3, p.786, 2009.

[5] S. Yu, X. Chen, W. Sun and D. Xie, A robust method for detecting and counting people, *Int. Conf. Audio, Language and Image Processing*, pp.1545-1549, 2008.

[6] T. H. Chen, T. Y. Chen and Z. X. Chen, An intelligent people-flow counting method for passing through a gate, *IEEE Conf. Robot. Autom. Mechatronics*, 2006.

[7] *http://www.people-sensing.com/more-features.*