International Journal of Innovative Computing, Information and Control
Volume **9**, Number **10**, October **2013**
 ISSN 1349-4198
pp. 4113–4129

A DISTANCE-BASED METHOD FOR COUNTING PASSERSBY USING SPACE-TIME IMAGERY

Ahmed Elmarhomy and Kenji Terada

Department of Information Science and Intelligent Systems
University of Tokushima
2-1 Minami Josanjima, Tokushima 770-8506, Japan
el_marhomy@iss.tokushima-u.ac.jp

Received September 2012; revised January 2013

Abstract. *Counting people and observing their movements are essential processes for security, organizational administration, and the improvement of corporate business results. This paper presents a method to automatically count passersby using virtual, vertical measurement lines. Recognizing passersby and generating space-time images are performed using an image sequence obtained from a USB camera. The space-time imagery represents recognizable regions surrounding each person which are identified through labeling-based segmentation. In this study, a space-time image represents time as pixel-distance, which an algorithm uses to achieve highly accurate counting. Human regions are based on space colors processed through template matching to select the optimal match in order to obtain direction and speed. Whether the shape represents one or two passersby is determined through the relationship between the passerby's speed and the human-pixel area. To evaluate the effectiveness of this proposed method, we performed experiments using a side-view camera placed near an entrance. The experiment achieved 97% accuracy in counting and direction.*
Keywords: Counting passersby, Counting people, Measurement lines, Space-time images, Pixel-distance, Direction, Speed, Template matching

1. **Introduction.** Security and security-related topics have recently received much attention, particularly national, personal, Internet, information, and banking security. Abraham Maslow's Hierarchy of Needs Theory provides a possible explanation for the abiding interest in security matters [1]. According to Maslow, after the most basic physiological needs are satisfied (the need for air, food and water, for example), human attention shifts to concerns of safety and security: protection from the natural elements, order, law and stability.

Within the multi-faceted field of security, many scholars have investigated video surveillance technology. These technologies have a wide range of applications in both restricted and open areas. Surveillance systems can provide extra protection for families, homes, and businesses [2,3]. Advances in object tracking have made it possible to obtain the spatiotemporal motion trajectories of moving objects from videos and use that data to analyze complex events. Moving-object tracking includes accurately counting people, an important task performed by automatic surveillance systems.

According to the Japanese National Police Agency (NPA), many types of violent crimes have increased and become serious problems for numerous institutions and commercial areas [4]. Therefore, observing people's movements for protection and security has become important for these commercial areas and companies, and counting people supports this goal [5]. Moreover, accurately counting people improves business results by accurately analyzing its performance. As management consultant Peter Drucker said, "If you cannot measure it, you cannot manage it." [6]. However, many such measurements are still performed manually [7]. Therefore, it is necessary to develop an automatic method to count passing people.

Some researchers have ventured proposals in this area [5,8]. Terada and Atsuta [5] proposed a method to automatically generate images of passersby from video recorded by Internet cameras; the images can be used later. One record (image) is generated for each person who comes into the camera's view. Portions of video not containing humans are skipped. Multiple frames capture images of a single passerby, but only the single frame offering the best image for storage is selected, while the rest are deleted. The proposed method assumed the use of Internet cameras rather than NTSC cameras. With the fast spread of the Internet, such cameras have become inexpensive and easy to install. Various types of cameras are used now for diverse purposes. Fujimoto and Terada [8] then proposed a method to count passersby with a camera. When counting passerby with a camera, individual passersby's movement must be tracked, but noise prevents accurate extraction of people. This paper proposed to identify and count people by using time-space images showing the length and breadth measurement lines [8].

Hashimoto et al. solved the problem of people counting by designing a specialized imaging system consisting of infrared (IR) sensitive cameras, mechanical chopping parts, and IR-transparent lenses [9]. The system uses background subtraction to create thermal images, placing more or less importance on different areas of the image depending on the region of interest, and analyzes them within a second. This array-based system can count people with accuracy of 95% under certain conditions. To work well, the system requires a distance of at least 10 cm to distinguish between passersby and thus to count them as two separate persons. The system also experiences problems of counting large movements by arms and legs. Therefore, this system is suitable for commercial centers with high-density traffic of people entering or exiting. In fact, people often go to supermarkets with their families, creating a close group of people that is the most difficult problem for counting-people systems to resolve.

Terada et al. created a system that can determine people's direction of movement and count them as they cross a virtual line [10]. This method avoids the problem of occlusion when groups of people pass through the camera's field of view. It uses a space-time image to determine the direction of people. Just like Terada et al. and Bahadori et al. employed stereo-vision in people tracking [11]. Their system uses continuous tracking and detection to handle the problem of people occlusion. Template-based tracking can stop detecting people as they become occluded, while eliminating false positives in tracking. Using multiple cameras improves the solution to the problem of occlusion, but for three-dimensional (3D) reconstruction, the two cameras must be calibrated well.

Haritaoglu and Flickner adopted another method to solve the problem of real-time tracking of people [12]. To segment silhouettes from the background, they used background subtraction based on the color and intensity of the pixel values. That information provides the basis to assign all the pixels in the image one of three classifications: foreground, background, and shadow. All the pixels classified as foreground are further divided into different regions. These foreground groups are then segmented into individual people using 2 different temporal and global motion constraints. To track these individuals, the algorithm uses an appearance model based on color and edge densities.

This paper proposes a method to automatically count passersby by recording images using virtual, vertical measurement lines. The process of recognizing a passerby is performed using an image sequence obtained from a Universal Serial Bus (USB) camera placed in a side-view position. While different types of cameras work from three different viewpoints (overhead, front, and side views) [14,15], the earlier proposed methods were not applicable to the widely installed side-view cameras selected for this work. This new approach uses a side-view camera that solves three new challenges: (1) two passersby walking in close proximity to each other, at the same time, and in the same direction; (2) two passersby moving simultaneously in opposite directions; and (3) a passerby moving in a line followed by another, or more, in quick succession.

In this study, space-time images represent time as pixel-distance, which an algorithm uses to achieve accurate counting. The passerby segmentation process detects the human regions. Template matching uses the RGB, YUV, and YIQ space colors to automatically select the best matches to determine the direction and speed of passersby. In the experiment, the camera was installed on the left side of a room near the entrance. A personal computer (PC) was connected to a camera, which had a rate of 17 frames per second. The results of the experiment confirm the effectiveness of this proposed method. It correctly detected every person and generated the passerby record image. Therefore, the desired goal – to detect and count the people passing by the camera – was successfully achieved.

2. **The Proposed Algorithm.** The proposed algorithm is shown in the flowchart in Figure 1. The first step is to acquire the frame from the camera. The next step is preprocessing, which includes background subtraction, removing noise, and establishing the four measurement lines. Next, movement detection serves as the trigger-point for the algorithm to begin generating space-time images. The human pixel area is then segmented, and template matching performed to determine direction and speed. Finally, counting is done based on the direction and number of passing people. The following sub-sections discuss each step of the algorithm in more detail.

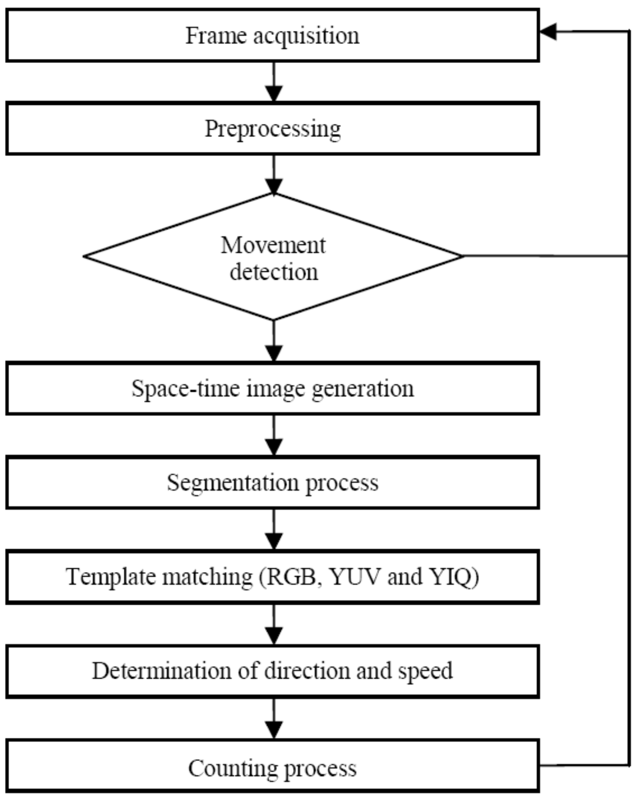


FIGURE 1. Algorithm stages

2.1. **Preprocessing.** A passerby is extracted from the frames in the preprocessing stage. The algorithm detects motion in three steps. First, in frame acquisition, a surveillance camera continuously captures and stores frames. Image data of passersby, from the time they enter the frame until they exit it, are extracted from the acquired series of images.

In the second step, a static image to be used as a static background image is constructed [16]. This image is acquired by capturing a frame without any motion. After subtracting the background reference from the captured frames, the remaining pixels represent possible motion. This subtraction process is repeated continuously. In the final step of preprocessing, noise is removed by applying a labeling process, which removes any irrelevant areas smaller than 200 pixels. This process reduces the noise caused by lighting and the color of clothing.

2.2. **Measurement line characteristics.** To represent and establish the measurement lines, four vertical lines, each 2 pixels wide, are set in the image. Whenever the line is wide, the passerby appears to be wide inside the space-time image, and the magnitude of the human-pixel area is represented with a large amount of pixels. As show in Figure 2, the two lines called the middle lines are in the middle of the image. The two remaining lines called the outer lines are to the left and right of the middle lines. The lines are positioned precisely to clearly determine the direction of movement. A separate background from the static background image is prepared for each of the four measurement lines. The data contained inside the four 2-pixel-wide measurement lines is used to generate space-time images.

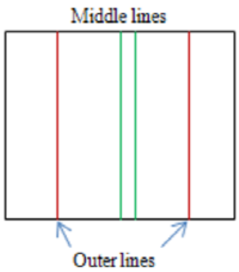


FIGURE 2. The measurement lines used to generate the space-time image

2.3. **Generating space-time images.** Background subtraction extracts human regions from the captured images, while a labeling filter removes the noise. Virtual measurement lines are superimposed on the original frame to obtain the measurement line images. Space-time images are produced by repeating this process and arranging measurement line images together with the x-axis (time) and the y-axis (space).

Figure 3 gives an example of how space-time images are generated. If motion is detected after subtracting the static background from a frame, the measurement lines of the image are captured. The corresponding static background measurement lines from the preprocessing stage are likewise continuously subtracted. The resulting difference in the subtraction process is continuously recorded in the space-time image.

The space-time image contains data for all passersby. When a passerby moves left or right, the resulting image appears, as shown in Figure 4. Since the measurement lines are vertical, the passersby are seen moving horizontally through the measurement lines, causing the passersby's shapes to appear vertically in the space-time image. Labeling is then applied to the space-time image to extract shapes of more than 200 pixels. In addition, because the x-axis represents time, the elapsed time can be calculated according to the number of frames.

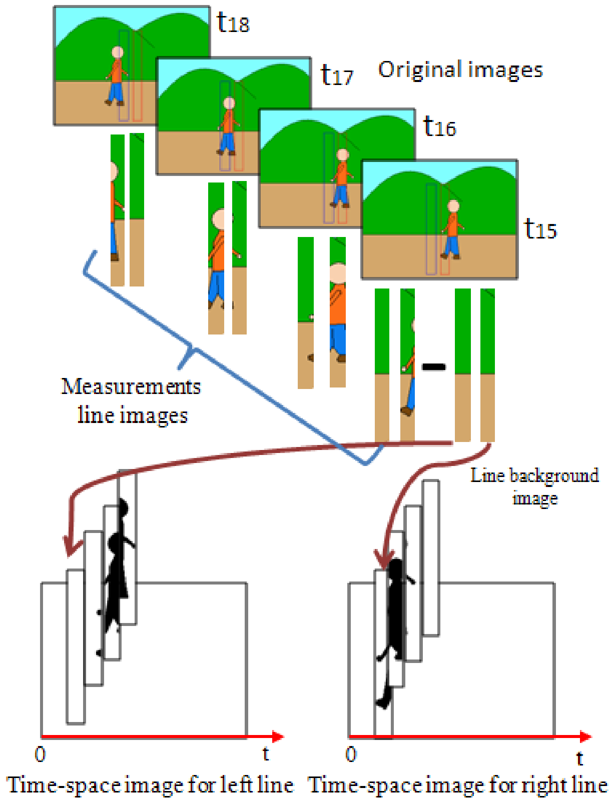


FIGURE 3. Space-time images generation according to time

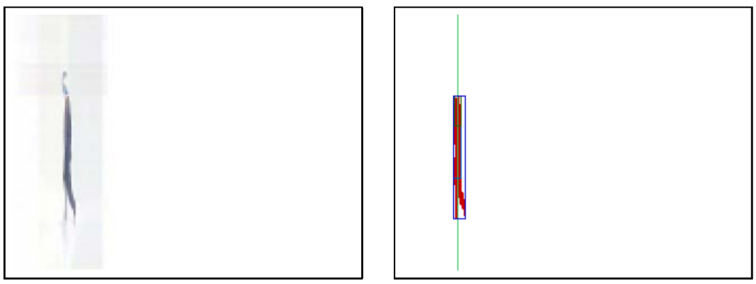


FIGURE 4. An example of the resulting space-time image

After generating the space-time image, the system treats the passerby as a single component, without specifying individual body parts. Therefore, a segmentation process is applied to the passerby, using labeling to assemble the disconnected components into a recognizable human shape.

2.4. **Segmentation of passersby.** The main goal of segmentation is to simplify or change the representation of an image into meaningful, more easily analyzed units [17]. Image segmentation locates objects and boundaries such as lines and curves in images. In this process, labels are assigned to all the pixels in an image that share certain visual characteristics [18,19]. Image segmentation creates a set of segments that together cover the entire image. All the pixels in each region possess a similar characteristic or computed property, such as color, intensity, or texture. Adjacent regions have significantly different characteristics [17].

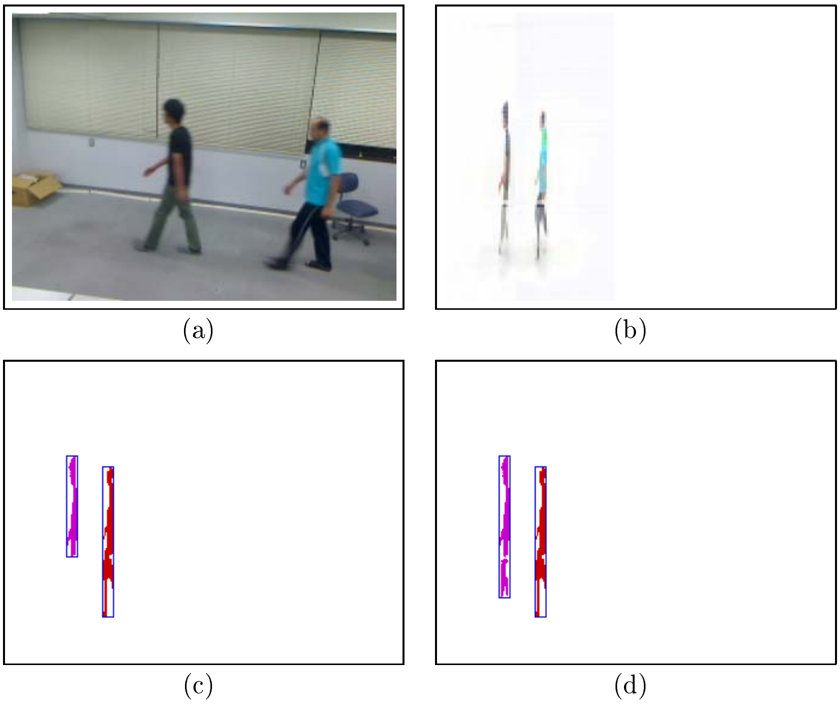


FIGURE 5. A passerby's shape before and after segmentation: (a) the original image; (b) the original color space-time image; (c) before segmentation; (d) after segmentation

One difficulty in segmentation is that the background noise sometimes produces different quantities of connected components for the same passerby. In the space-time image, the shape of each passerby appears almost identical, so segmentation is sometimes necessary to assemble the appropriate shape of each passerby. To solve this problem, the system calculates and counts the connected components that represent the same passerby with different labels and then assigns all the connected components the same labeling number so that the passerby is represented as one component. The passerby's position is vertical, so the system vertically searches for the other connected components of the passerby between the left and right boundaries of the pre-segmented connected components. Figure 5 shows an example of a passerby's shape in a space-time image before and after segmentation. This problem also influences the template matching process and affects the magnitude and size of the human-pixel area, which is discussed in more detail later.

2.5. **Template matching.** In general, template matching uses binary images. For our proposed algorithm, accurately matching more than one passerby in the same space-time image is difficult because of the problem of mismatching [17]. For example, passersby's shapes are nearly identical when using binary images, contributing to the problem of mismatching. When dealing with more than one passerby, space colors such as RGB, YUV and YIQ produce more accurate matching [19,20], using the following formula:

$$\left.\begin{aligned} Y &= 0.229 \times R + 0.587 \times G + 0.114 \times B \\ U &= -0.147 \times R - 0.289 \times G + 0.436 \times B + 128 \\ V &= 0.615 \times R - 0.515 \times G - 0.100 \times B + 128 \end{aligned}\right\} \quad (1)$$

$$\left.\begin{array}{l} Y = 0.229 \times R + 0.587 \times G + 0.114 \times B \\ I = 0.596 \times R - 0.275 \times G - 0.321 \times B + 128 \\ Q = 0.212 \times R - 0.532 \times G + 0.311 \times B + 128 \end{array}\right\} \quad (2)$$

2.6. **Optimal match.** [Note: This section introduces two acronyms: MPD for measured pixel-distance and TPD for template-matching-resulting pixel-distance.] A match can be achieved by treating the area of a passerby as a template image taken from the left middle-measurement line of the space-time image and then performing template matching. When two shapes are detected in the space-time image (for an example, see Figure 2), the pixel-distance between the two shapes is measured (MPD). The template matching process uses different space colors (RGB or YUV), as shown in Figure 6, and the result covers the whole, or part of, the passerby's shape. The label of the shape can then be identified by reviewing the labeling lookup table and detecting old labels inside the matching result area. After determining the labels of the two shapes, the pixel-distance between the two resulting shapes in the same space-time image is likewise measured (TPD). Comparing the MPD and TPD then yields the optimal match.

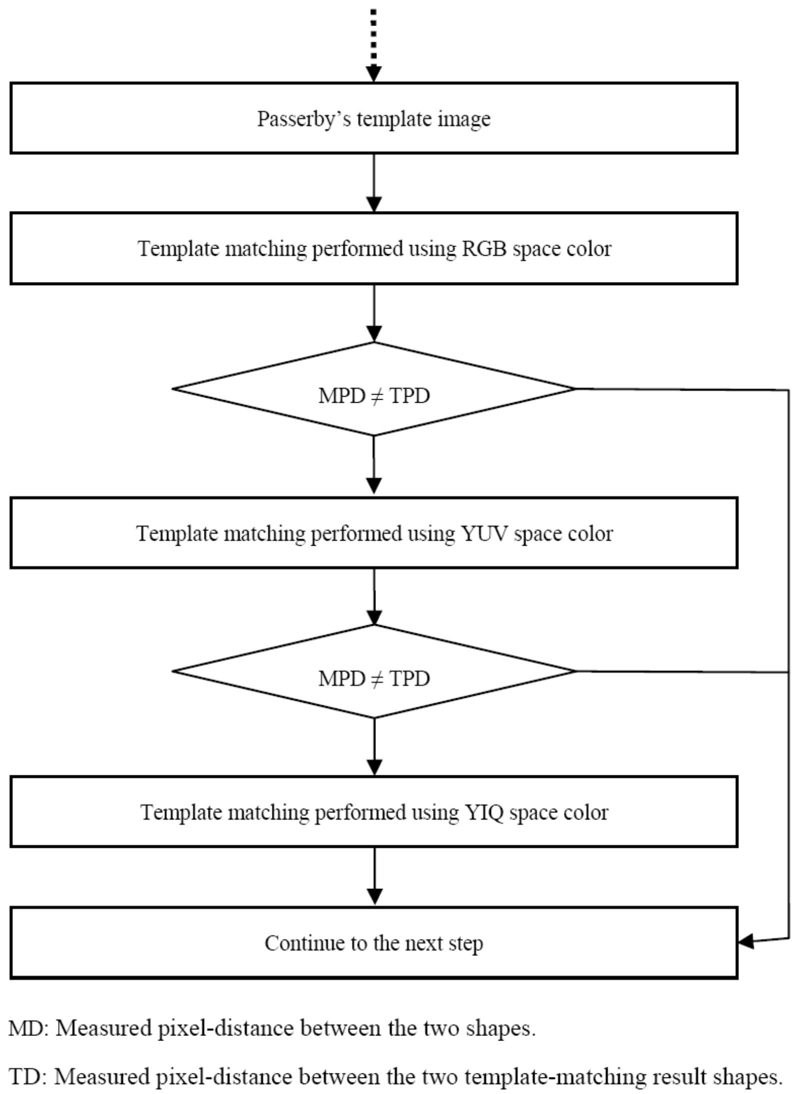


FIGURE 6. Optimal match

2.7. **Determination of direction and speed.** This section discusses how measurement lines are used to determine the speed and direction of passersby.

2.7.1. *Detection of the direction of passersby.* To determine the direction of passersby, two space-time images, one for each middle measurement line, are used. The distance between the two middle measurement lines needs to be considered. A passerby completely crosses one middle measurement line before crossing the second. The distance between the middle measurement lines is sufficient to detect the direction of at least one normal walking step. Therefore, measuring the difference between the passerby's position in the two space-time images can determine the direction of motion.

After template matching, a match can be achieved. The passerby's exact position can be determined by applying the labeling concepts in the reference table to the resulting match. Comparing the left position of the passerby in both space-time images shows the person's direction, as illustrated in Figure 7. Using the passerby's exact position to detect the direction produces more accurate results than using the passerby's matching position.

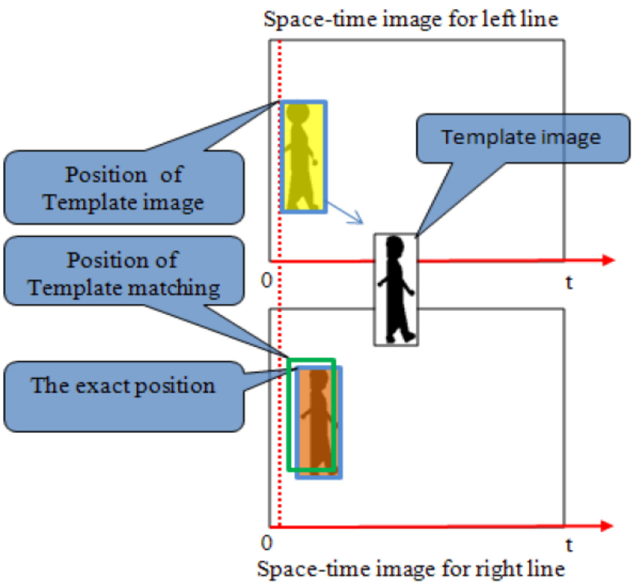


Figure 7. Space-time image for direction detection

2.7.2. *Measurement of the speed of passerby.* To determine speed, the two middle measurement lines space-time images are used. The distance between them is measured manually in centimeters, as shown in Figure 8(a). The calculation of the elapsed time is discussed later.

Head Position. Since the measurement lines are vertical, passersby move horizontally through the measurement lines, causing their shape to appear vertically in the space-time image, as discussed in Subsection 2.4. Dividing the passerby's shape into three equal parts and taking the uppermost part as the passerby detects the position of the head, as shown in Figures 8(b) and 8(c). Therefore, the position of the passerby's head is detected twice, once by each middle measurement lines in the space-time image.

Time Determination. The difference between the passerby's position in the two space-time images is used to determine the direction of the passerby. Since the x-axis represents time, this difference can be used to calculate the elapsed time. The calculated time, though, is not the precise elapsed time, which is needed; without the correct elapsed time, the speed calculation will be inaccurate. In this case, the exact elapsed time can

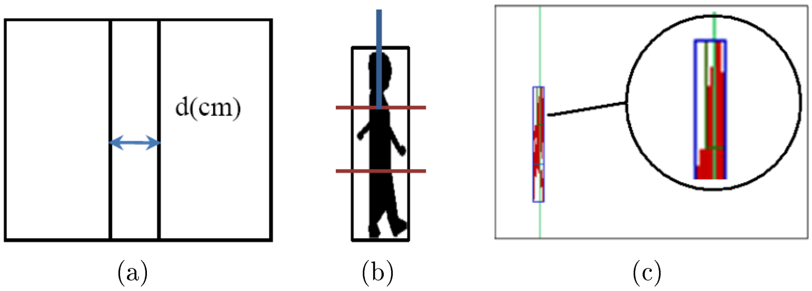


FIGURE 8. Speed calculation with (a) the distance in d(cm); (b) and (c) represent the passerby's head.

be calculated by using the vertical center of the body position. The center of the body position can be calculated using the head position. The difference between the centers of the passerby's body position in the two space-time images is calculated in pixels. Since each measurement line is 2 pixels wide, the number of frames can be counted. As a result, multiplying the frame rate by the number of frames yields the precise elapsed time.

Speed Calculation. After calculating the precise elapsed time, dividing the distance by the elapsed time yields the passerby's speed.

2.8. **Human-pixel area.** The human-pixel area is the number of pixels that represent the magnitude of the passerby's shape in the space-time image. Four important factors influence the size and magnitude of the passerby's shape (human-pixel area): the measurement line width, passerby segmentation, frame rate, and speed of the passerby. Subsections 2.3 and 2.4 discuss passerby segmentation and the width of the measurement lines. Using different examples, the following discussion focuses on the influence of the frame rate and passerby's speed.

- Frame rate: As illustrated in the following cases, the influence of the frame rate on the human-pixel area can be observed. When using a slow frame rate to acquire the frame, the passerby appears to be thin inside the space-time image, and a small amount of pixels represents the magnitude of the human-pixel area. On the other hand, when using a high frame rate to acquire the frame, the passerby appears to be wide, and a large amount of pixels represents the magnitude of the human-pixel area.
- Speed: When the frame rate is constant, the influence of the passerby's speed on the human-pixel area can be clearly observed. When a passerby's speed is high (fast), the passerby appears thin inside the space-time image, and the magnitude of the human-pixel area is represented with a smaller amount of pixels than that used to represent a passerby walking at a normal speed. On the other hand, when a passerby's speed is slow, the passerby appears to be wide inside the space-time image, and the magnitude of the human-pixel area is represented with a larger amount of pixels than that used to represent a passerby walking at a normal speed.

2.9. **Pixel-speed ratio.** The human-pixel area is counted, and then the passerby's speed is determined. Multiplying the human-pixel area by the speed of the passerby generates the pixel-speed ratio (R), as shown in Equation (3):

$$R = \textit{human-pixel area} \times \textit{passerby speed} \tag{3}$$

2.10. **Counting process.** This section discusses in detail the counting process for different situations: (1) one or two passersby moving in the same direction, (2) two passersby moving in opposite directions, and (3) one passerby followed by another. Figure 9 provides an overview of the counting process algorithm. As noted earlier, the direction, speed, and human-pixels area are determined first, and then each passerby is counted based on the direction of movement. The following sections explain the counting process for each situation.

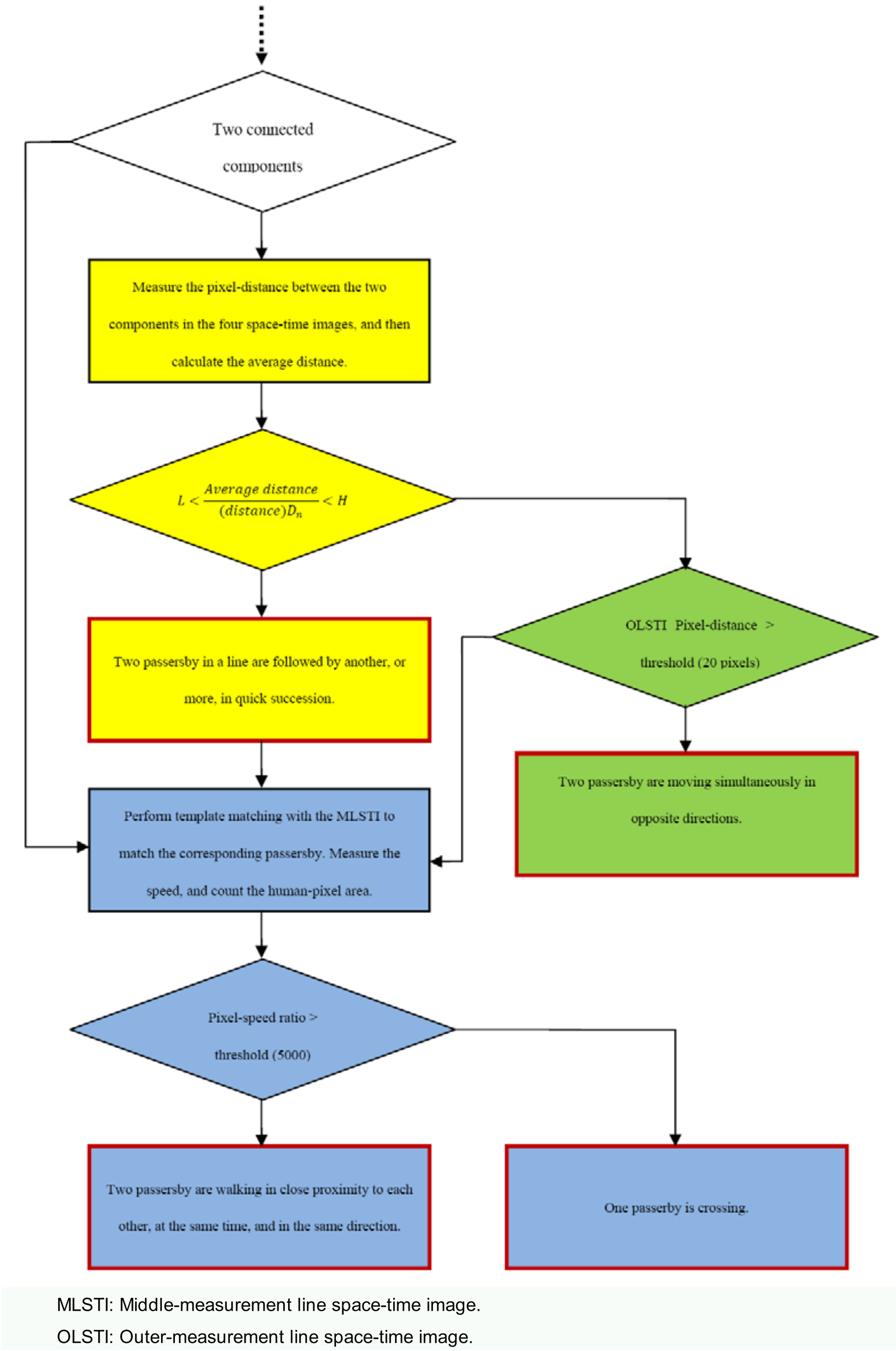


FIGURE 9. Counting processing algorithm

2.10.1. *Counting passersby walking in the same direction.* When two passersby walk in close proximity, at the same time, and in the same direction, their combined shape appears to be wide in the space-time image, and the magnitude of human-pixel area is represented with a large amount of pixels. This situation is similar to that of one passerby passing at a slow speed, as discussed in Subsection 2.8. Therefore, using only the passerby's shape, it is difficult to determine whether there are one or two passersby. Therefore, a ratio to distinguish between single and multiple shapes is needed.

Counting Using the Pixel-speed Ratio. The pixel-speed ratio can determine wheth-er the shape is of one or two passersby. If the average value of the ratio in the two middle-measurement line space-time images is more than the value of the threshold chosen after many experiments (set at 5,000 in this work), the system detects two passersby walking in the same direction; otherwise, the system detects one passerby. In other words,

$$\text{if } R \begin{cases} < \text{threshold} & \textit{one passerby} \\ > \text{threshold} & \textit{two passerby} \end{cases}$$

2.10.2. *Counting passersby walking in opposite directions.* Instead of the two middle measurement lines, the two outer measurement lines in the two space-time images are used to count two passersby walking in opposite directions. A passerby crosses one of the two outer lines before crossing the second one. The time needed to reach the second outer line is represented with distance in the space-time image. If two connected components are detected in one or both of the space-time images, the distance between the two passersby is measured in pixels. If the pixel-distance is greater than the established threshold for the elapsed time to complete crossing the two outer lines (set at 20 pixels in this work), the system recognizes and counts two passersby walking in opposite directions.

2.10.3. *Counting one passerby followed by others.* The counting function is modified to count passersby following each other in a line in quick succession. When two connected component shapes are detected, the pixel-distance (D) between the shapes in the space-time images is measured. The measurement process is repeated and applied to all four space-time images. The average value of the four distances is then calculated using the following Equation (4):

$$Average\ distance = \frac{D_1 + D_2 + D_3 + D_4}{4} \tag{4}$$

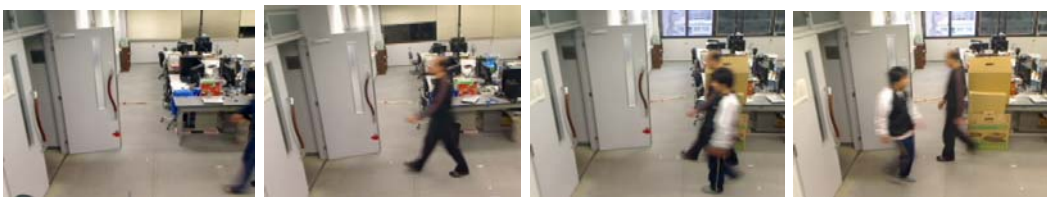
After calculating the average value, the relationship between the average value and the distance is defined based on Equation (5). The main purpose of this equation is to determine the proportional (commensurate) of the four distances. The low (L) and high (H) values are threshold values. After experimentation, the most effective L and H values for the purposes of counting are chosen. Using Equation (5), if the four values of the dividing results have achieved a relationship, the system detects the shapes of two passersby followed by one another.

$$L < \frac{Average\ distance}{(distance)D_n} < H \tag{5}$$

Finally, template matching is performed to match the corresponding passersby. Therefore, achieving optimal matching requires using the measured pixel-distance between the passersby in the space-time images. In addition, the pixel-speed ratio is calculated to determine whether the shape is of one or two passersby. This processing is done independently for each passerby.

3. Experiments and Results.

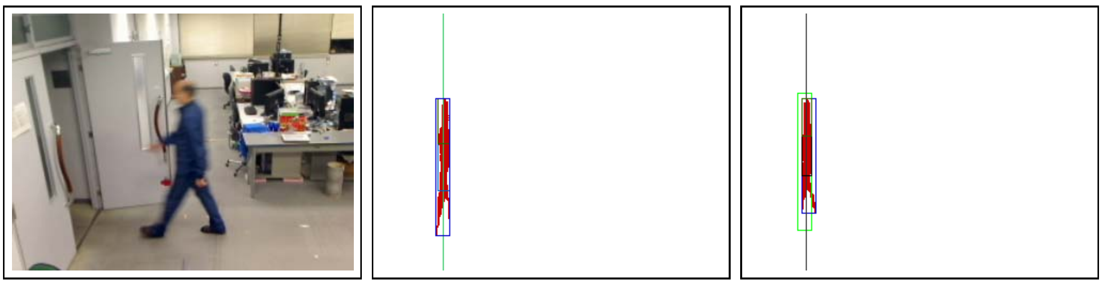
3.1. Experimental data. The method was tested using a series of video sequences of various scenarios. Figure 10 shows images captured by a USB camera installed on the left side of a room near the entrance which captured 320×240 pixel images at an average of 17 frames per second. In Figure 10(a), a person enters the frame zone, and the algorithm detects motion. In Figure 10(b), one passerby crosses in front of the camera. In Figure 10(c), two passersby walk in close proximity to each other, at the same time, and in the same direction. In Figure 10(d), two passersby move simultaneously in opposite directions.



(a) (b) (c) (d)

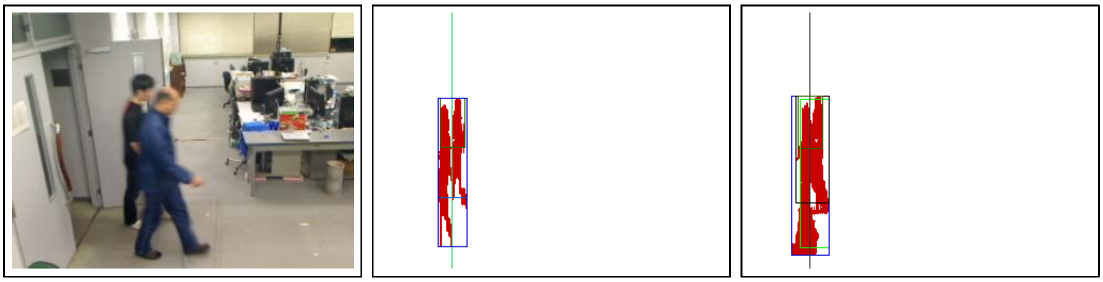
Figure 10. Frames acquired using the surveillance camera: (a) and (b) show a single person passing, (c) and (d) show different examples of two people passing.

3.2. Experimental observations. The method successfully counted a single passerby walking in any direction, incoming or outgoing, based on the pixel-speed ratio using the middle-measurement lines space-time images, as shown in Figure 11. When two passersby walked together at the same time in the same direction, the method counted the two passersby based on the pixel-speed ratio using the middle-measurement line space-time images, as shown in Figure 12. When two passersby walked in opposite directions, the method precisely counted the two passersby based on the measured pixel-distance between them in the outer-measurement line space-time image, as shown in Figure 13. Finally, multiple passersby in a line following one another in quick succession were counted based on the measured pixel-distance between two passersby in the four space-time images (Equations (4) and (5)) and the pixel-speed ratio, as shown in Figure 11. In Figure 14, in which two passersby follow one another, the method measured the pixel-distance between the two shapes found in all four space-time images.



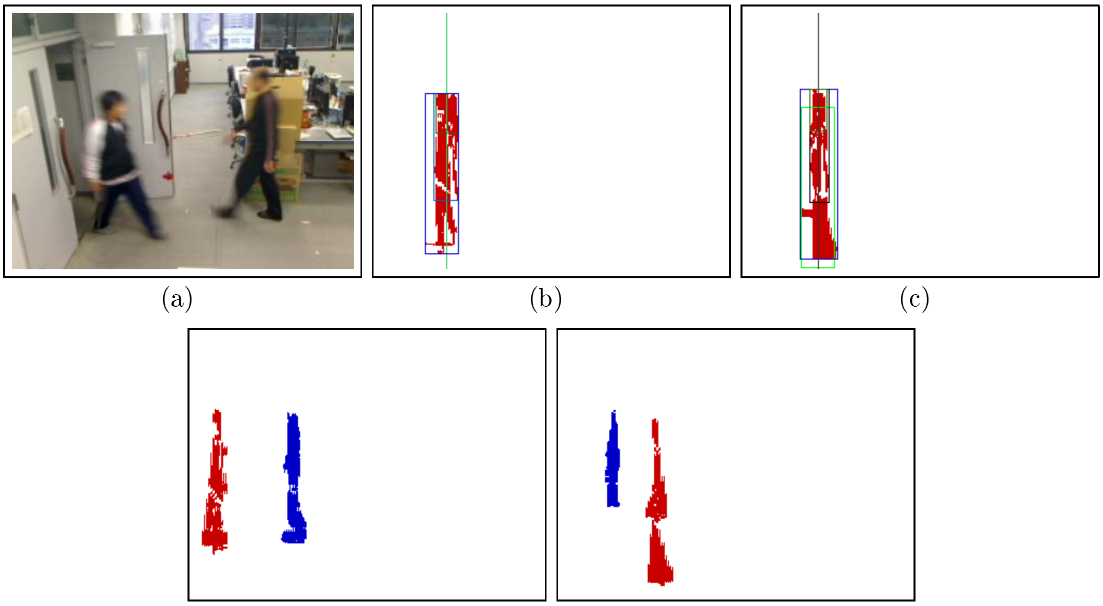
(a) (b) (c)

Figure 11. Single passerby walking in the direction of the exit: (a) the original images, (b) and (c) the left and right middle-measurement lines space-time images



(a) (b) (c)

FIGURE 12. Two passersby walking together in the same direction: (a) the original images, (b) and (c) the left and right middle-measurement lines space-time images



(a) (b) (c)

(d) (e)

FIGURE 13. Two passersby walking in opposite directions: (a) the original images, (b) and (c) the left and right middle-measurement lines space-time images, and (d) and (e) the outer measurement-lines space-time images

3.3. **Experimental results.** This section discusses our experimental results which demonstrate the successful, accurate matching and automatic counting of passersby in various cases and directions in different video sequences.

3.3.1. *Accurate matching.* Accurately matching more than one passerby in the same space-time image is difficult because of the problem of mismatching. When dealing with more than one passerby, the system uses space colors such as RGB, YUV and YIQ to achieve accurate matching. When the method used only RGB space colors in template matching, the error rate of the results was approximately 15% to 25%. Using RGB and YUV space colors reduced the error rate to approximately 7% to 12%. Finally, when using three space colors (RGB, YUV and YIQ), the error rate was virtually unnoticeable (approximately 3%). Figure 15 illustrates the determination of optimal matching.

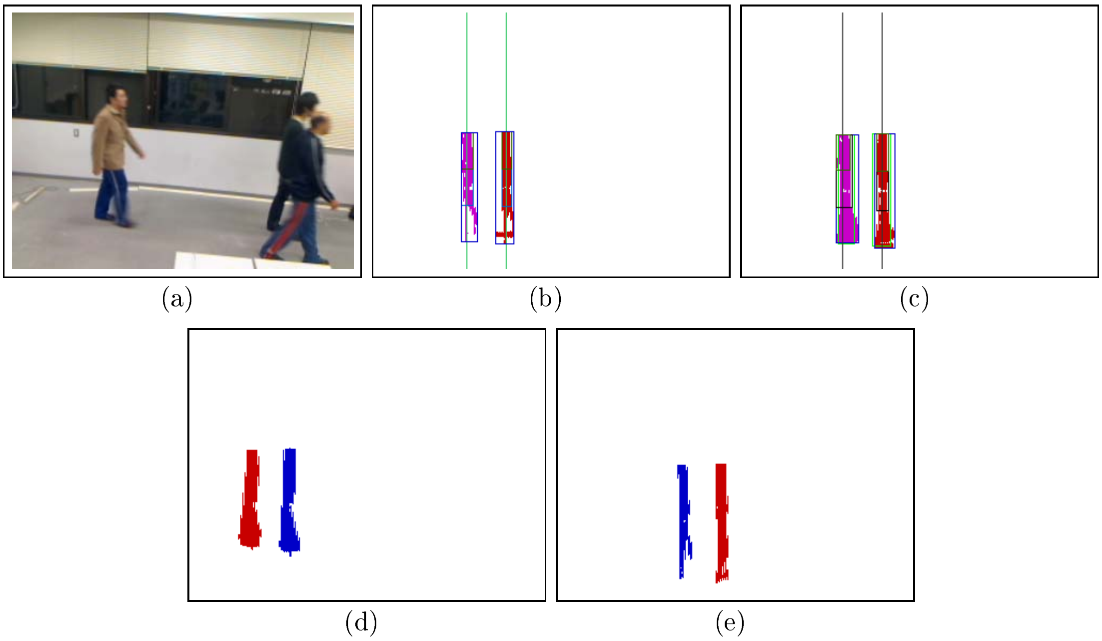


FIGURE 14. Two passersby walking in close proximity followed by another: (a) the original images, (b) and (c) the left and right middle-measurement lines space-time images, and (d) and (e) the outer measurement-lines space-time images

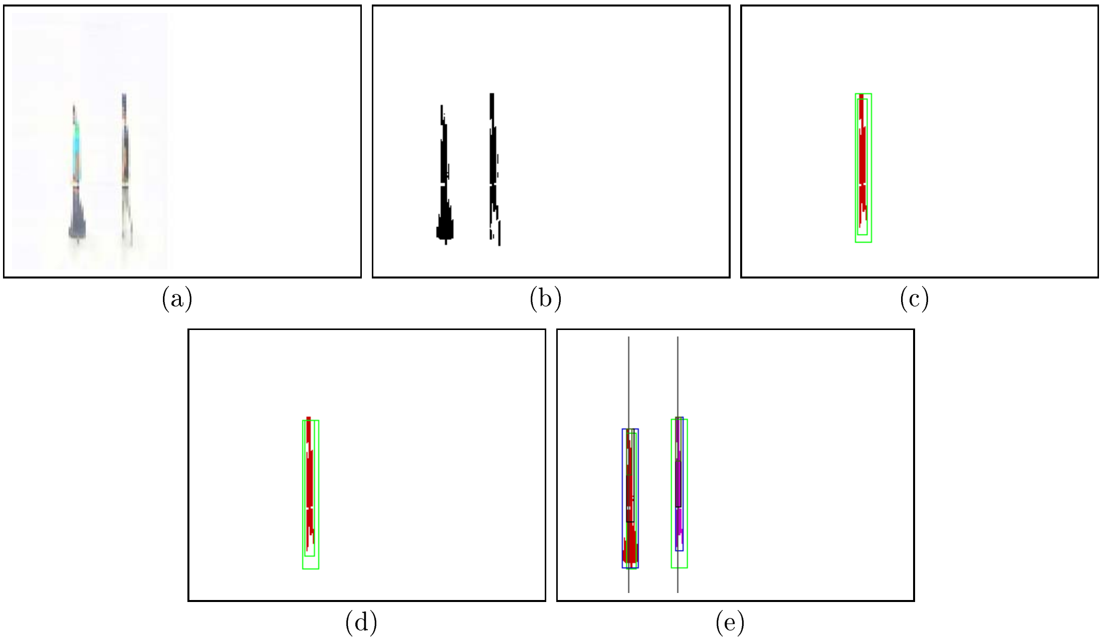


FIGURE 15. Template matching that determines the optimal match: (a) color space-time image, (b) binary image, (c) and (d) mismatching with RGB and YUV space colors, (e) correct matching with YIQ space colors

3.3.2. *Passerby counting accuracy.* The method automatically counted the passersby in various cases and directions from different video sequences. Table 1 shows the counting accuracy for multiple experiments with different situations: one or two passersby moving in the same direction, two passersby moving in opposite directions, and one passerby

Table 1. Experimental results of the counting algorithm in various situations

Status		Number of passersby	Undetected	Detected direction (%)	Speed	Speed-pixel ratio	**Counting accuracy (%)**
One passerby		One	0	100	Measured	Used	**100**
Two in close proximity		Two	0	100	Measured	Used	**90**
Opposite direction		Two	0	100	Not measured	Not used	**100**
Passersby in a line followed by another, or more	One followed by one	Two	0	100	Measured for each	Used	**100**
	One followed by two in close proximity	Three	0	100	Measured for each	Used	**95**
	Two in close proximity followed by two in close proximity	Four	0	100	Measured for each	Used	**90**

Table 2. Accuracy of the system's counts of people passing the camera in a fixed time

Time (min)	Manual count		System count		Accuracy (%)
	Exit	Enter	Exit	Enter	
5	18	21	18	20	99
5	15	15	15	15	100
5	11	18	12	18	98
5	16	17	16	17	100

followed by another when the number of passersby is one, two, three or four and the speed is measured only sometimes. The results for 50 cases in each situation, presented in Table 1, confirm that the new method effectively and efficiently counts passersby.

Table 2 displays a sample of the system's accuracy at counting passersby in a fixed time (5 minutes). From the sample results of the long experiment, the manual count was Exit: 185 and Enter: 209, and the method determined Exit: 180 and Enter: 205. Significantly, the number of passersby was determined and counted successfully, with a high accuracy of approximately 97%.

3.4. **Discussion.** This work used five characteristics: the position of a person's head, the center of gravity, human-pixel area, speed of passerby, and distance between people. These five characteristics enable accurate counting of passersby. The proposed method does not involve optical flow or other algorithms at this level. Instead, human images are extracted and tracked using background subtraction and time-space images. Our method used the widely installed side-view camera, for which the earlier approaches were not applicable. On the other hand, the overhead camera is useful for only one purpose – counting people – and cannot be used for any other functions compared with the side-view camera. Our method does not require the same conditions as the earlier methods, such as a distance of at least 10 cm to distinguish passersby and thus count them as two separate people. Our method can overcome this challenge and, in this case, count two passersby using the five previously mentioned characteristics. In addition, the earlier methods sometimes failed to count people with large arm and leg movements. This proposed method, however, can count not only one passerby but also two passersby walking in close proximity at the same time in the same direction or opposite directions and passersby moving in a line in quick succession. As mentioned earlier, using different space colors to perform template matching and automatically select the optimal matching accurately counts passersby with an error rate of approximately 3%, lower than earlier proposed methods.

4. **Conclusion.** This paper proposes a new approach to automatically count passersby using four virtual, vertical measurement lines, which are precisely positioned in a frame

to clearly determine the direction of movement. Four space-time images are generated, one for each measurement line, based on a sequence of images obtained from a USB camera. The space-time images represent human regions, which are treated with labeling to remove any noise.

In the proposed method, accurately matching more than one passerby in the same space-time image is difficult because of the problem of mismatching. Overcoming mismatching requires assembling the correct shape of each passerby through the segmentation process so that the shape appears as one component. Additionally, different space colors are used to perform template matching, and then the system automatically selects the optimal matching (the error rate of the results is approximately 3%) to determine direction and speed.

In the experiments, a side-view camera was fixed on the left side of the room near the entrance. A PC was connected to the camera, which had a rate of 17 frames per second. The experiment results confirm that the new method effectively and efficiently counts passersby. The method was tested in multiple situations: (1) one passerby walking in any direction; (2) two passersby walking in close proximity, at the same time, and in the same direction; (3) two passersby moving simultaneously in opposite directions; and (4) a passerby followed by another, or more, in a line in quick succession. As an additional, significant result, the number of passerby was determined and counted successfully. This accurate counting used the pixel-distance between passersby and the relationship between the speed of the passersby and the human-pixel area in the space-time images.

Future work could focus on making the background subtraction process more sensitive to environmental changes and automating updating of the background. Additional improvements could include enabling the method to count groups of people.

REFERENCES

[1] R. Cianci and P. A. Gambrel, Maslow's hierarchy of needs: Does it apply in a collectivist culture? *Journal of Applied Management and Entrepreneurship*, vol.8, no.2, pp.143-161, 2003.

[2] A. Cavoukian, *Guidelines for Using Video Surveillance Cameras in Public Places*, Information and Privacy Commissioner, Ontario, 2001.

[3] M. Gray, Urban surveillance and panopticism: Will we recognize the facial recognition society? *Surveillance and Society*, vol.1, no.3, pp.314-330, 2003.

[4] *NPA White Paper*, http://www.npa.gov.jp/hakusho.

[5] K. Terada and K. Atsuta, Automatic generation of passerby record images using Internet cameras, *Electronics and Communications in Japan*, vol.92, no.11, pp.553-560, 2009.

[6] J. A. Byrne and L. Gerdes, The man who invented management, *BusinessWeek*, 2009.

[7] K. Terada, D. Yoshida, S. Oe and J. Yamaguchi, A counting method of the number of passing people using a stereo camera, *IEEE Proc. of Industrial Electronics Conference*, vol.3, pp.1318-1323, 1999.

[8] K. Fujimoto and K. Terada, Method for counting passerby by length and breadth measurement line, *Proc. of the 16th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, Hiroshima, pp.436-439, 2010.

[9] K. Hashimoto, K. Morinaka, N. Yoshiike, C. Kawaguchi and S. Matsueda, People count system using multi-sensing application, *Proc. of the International Conference on Solid State Sensors and Actuators*, 1997.

[10] K. Terada, D. Yoshida, S. Oe and J. Yamaguchi, A method of counting the passing people by using the stereo images, *Proc. of the International Conference on Image Processing*, 1999.

[11] S. Bahadori, G. Grisetti, L. Iocchi, G. R. Leone and D. Nardi, Real-time tracking of multiple people through stereo vision, *Proc. of IEEE International Workshop on Intelligent Environments*, 2005.

[12] I. Haritaoglu and M. Flickner, Detection and tracking of shopping groups in stores, *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[13] K. Terada and K. Matsubara, A method for counting multidirection passer-by by using circular space-time image, *IEEJ Transactions on Electronics, Information and Systems*, vol.129, no.6, pp.1016-1025, 2009.

[14] A. Elmarhomy, S. Karungaru and K. Terada, A method for counting passersby using time-space image, *Proc. of the 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, 2011.
[15] A. Elmarhomy, S. Karungaru and K. Terada, A new approach for counting passersby utilizing space-time images, *International Conference on Digital Image and Multidimensional Signal Processing*, pp.1912-1919, 2012.
[16] T. Matsuyama, T. Wada, H. Habe and K. Tanahashi, Background subtraction under varying illumination, *Transactions on IEICE*, pp.2201-2211, 2001.
[17] G. L. Shapiro and C. G. Stockman, *Computer Vision*, Prentice-Hall, New Jersey, 2001.
[18] A. Yang, J. Wright, Y. Ma and S. Sastry, Unsupervised segmentation of natural images via lossy data compression, *Computer Vision and Image Understanding*, vol.110, no.2, pp.212-225, 2008.
[19] S. Yu, Segmentation induced by scale invariance, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
[20] R. Brunelli, *Template Matching Techniques in Computer Vision: Theory and Practice*, Wiley, 2009.
[21] J. Watkinson, *The Art of Digital Video*, Focal Press, 2008.
[22] SMPTE, Annotated glossary of essential terms for electronic production, *Engineering Guideline EG 28*, 1993.
[23] E. P. J. Tozer, *Broadcast Engineer's Reference Book*, Elsevier, 2004.