A Novel Layer-Scanning Method
for Improving Real-Time People Counting

Daw-Tung Lin and Dong-Han Jhuang

Department of Computer Science and Information Engineering,
No.151, University Rd., Sansia Dist., New Taipei City 23741, Taiwan (R.O.C.)
dalton@mail.ntpu.edu.tw, s49885046@webmail.ntpu.edu.tw

Abstract. People counting is an important subject for various applications and analyses. This paper proposes a novel method for counting the number of targets by using the layer scanning of depth information provided by the Kinect® sensor. The steps of this method include constructing a depth image background model, deriving foreground depth map, filtering the noise, classifying the targets, and screening the area of targets with layer scanning to calculate the number of targets by determining the highest position of the respective targets, tracking and analyzing the objects, and counting the number of the objects. Moreover, the dynamic number of targets is calculated using a tracking algorithm. The proposed system is beneficial in automatic, effective, and precise calculation of the number of targets in a specific area. Furthermore, the technique of the present method is not affected by the changes of the ambient light which can effectively reduce the interference of the background.

Keywords: People counting, Kinect®, depth information, layer scanning, object tracking.

1 Introduction

Calculating the number of the objects within a certain regional space is important to manage the objects. Although infrared rays and ultrasonic sensors have been used for people counting, no perfect solution can overcome crowded and occluded situations. Several methods have been proposed to count passersby in several indoor spaces with a mounted overhead camera. Masoud and Papanikolopoulos [1] implemented a blob-tracking method that can alleviate several occlusions or overlapping conditions. Chen [2] implemented a useful method on solving the merge-split problem when tracking people. The purpose of [3] and [4] was to solve the crowd-counting issue by using various methods. Chen et al. [5] expanded the field of counting people to a more applicable level. Antic et al. [6] used the K-means algorithm to analyze the number of passing people. Lu et al. [7] have applied the Kinect® sensor to track people. In order to overcome the disadvantage of the various instruments in the prior art, this work uses a Kinect® infrared depth sensor and supplemented by the detection method to provide a real-time layer scanning method to enhance the accuracy of the counting objects.

2   System Architecture

The proposed people-counting system uses a Kinect$^{®}$ sensor mounted on a
ceiling to allow the sensor to capture the downward-vertical viewing field.
Figure 1(a) shows the scenario of the system setup and application. The scene
has incoming and outgoing pedestrians, and they are detected and tracked when
they enter the viewing region. Figure 1(b) shows the complete flowchart of the
proposed procedure and implementation, which comprises three parts: depth im-
age background modeling, the layer-scanning method, and objects analysis and
counting. The first N depth frames are averaged for use as the main background
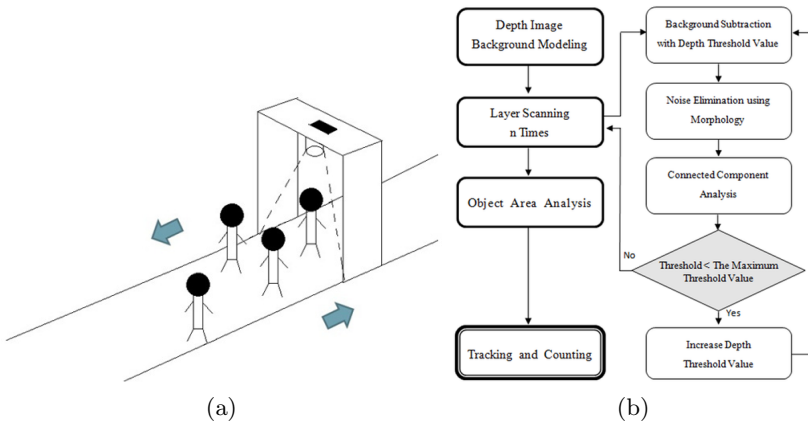for the image subtraction step in layer-scanning.

Fig. 1. (a) The people counting scenario; (b) The complete system flowchart

3   Proposed Layer Scanning Algorithm

3.1   Depth Background Subtraction and Value of Interest(VOI)
       Foreground Extraction

In contrast to the ordinary method of background subtraction, the image that is
analyzed in this study is a depth image. Background subtraction is employed to
derive a foreground object which means shooting a (N+1)th frame where objects
appears and comparing with the background map constructed in the previous N
frames as illustrated in Section 2. The background will be updated periodically.
After setting this value according to our interests (value of interest (VOI)),
the appropriate foreground was obtained after background subtraction. Next,
morphological opening and closing operations followed by connected component
analysis were applied to remove the noise and obtain a smoother foreground for
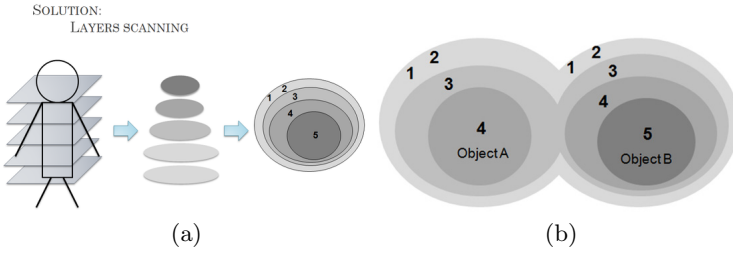post-estimation.

Fig. 2. (a) Layer-scanning concept diagram; (b) Cluttered objects separation using layer-scanning method with different depth range layer labels

3.2 Layer-Scanning Algorithm

Figure 2(a) shows the fundamental concept of this method. Because the direction of the proposed sensor is vertically downward, the shapes of the pedestrians in the scene are similar to ellipses. The area of projection of the upper portion of a person is normally smaller compared to the lower one; thus, if when comparing and stacking the areas of every layer from low to high, the top portion of each object can be estimated. Algorithm 1 shows the pseudo-code of the proposed layer scanning method.

Algorithm 1. Layer-scanning

1: **initialization** value of interest(VOI) depth threshold
2: **while** VOI threshold does not reach limit value **do**
3:     background subtraction
4:     morphology(opening and closing)
5:     connected component analysis
6:     **copy** processed object area and provide new label mark to the object area
7:     **if** VOI depth threshold is less than limit value **then**
8:         increase VOI depth threshold by one layer value
9:     **end if**
10: **end while**

3.3 Object Area Analysis and Cluttered Objects Separation

An area range condition was set to eliminate the smashed area based on $A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n = result$, where $low\_area\_value < A_i < high\_area\_value$, $0 < i \leq n$. Generally, a first loop scanned by the first scan is shown in Fig. 2(a) labeled by 1. The range 1 is the depth range of the knee to the foot of a human, the foreground depth map thereof has a lighter grayscale value. Superposing the obtained foreground depth maps respectively, and removing the discrepant area to obtain a highest height portion, thereby determining the highest position of the object is the range n. Furthermore, the cluttered objects can be correctly separated by identifying the above mentioned highest portion of the objects using different depth range layer-scanning. Therefore, the proposed method is superior to

the other state-of-the-art approaches in the aspect of solving the crowd-counting issue. Figure 2(b) depicts the concept of cluttered objects separation. The foreground objects are correctly identified as two objects using depth range labels 4 and 5, even-though these two objects are connected with each other.

3.4   Tracking Algorithm and Counting

After using the proposed layer scanning method, each frame may possess candidate moving objects. The objects adjacent to each other are reclassified to obtain the number of the objects. The tracking in this study may fulfilled by comparing the foreground depth map in consecutive video frames and selecting the shortest distance for pairing. Furthermore, the vector of each object can be estimated by calculating the difference in the coordinates of an object. Thereafter, the vector information can be used to identify whether an object is moving upward or downward.

4   Experimental Results and Discussion

The proposed real-time people-counting system was implemented to operate at approximately 30 FPS. The sensor is set approximately 3 m above the floor. All of the experiments were performed in real-time. Several conditions of the experiments are shown in Fig. 3, including normal speed walking, fast moving, bi-directional moving, several people walking closely, etc.. The proposed people-counting system can, on average, count people with an accuracy of 96.11% in the range of 50 to 200 cm. Table 1 presents the results of the experiments. Figure 3(c) shows the situation of false counting caused by an object of a pedestrian has the same depth value as the pedestrian. To solve this issue, the feature value
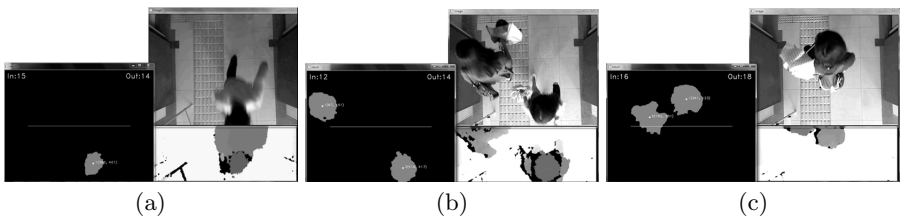

(a)   (b)   (c)

Fig. 3. (a) One fast moving person test; (b) Two people walking in bi-directional way; (c) False count condition

Table 1. Experimental results

Test	In	Out	Correct In	Correct Out	Accuracy
1	57	38	55	37	96.93%
2	103	125	99	123	97.25%
3	223	195	211	187	95.24%

regarding shape can be integrated into the system or the color histogram can be used to determine more accurately if an object is human. Moreover, the fast-moving objects and sudden switching on and off of the light test indicated that both of these conditions do not influence the people-counting result.

5 Conclusions

This paper proposed a novel method for counting the number of the targets using the layer scanning method based on the depth information provided by Kinect$^{®}$ sensor. The steps of this method includes constructing a depth image background model, filtering the noise of foreground frame and classifying the targets, and screening the area of targets based on layer scanning to calculate the number of targets by determining the highest positions of the respective targets. The depth data and proposed algorithm were used to view the top region of each object with the Kinect$^{®}$ sensor; thus, each object was tracked correctly, despite their lower portions being connected and the highest portions not being connected. In addition, the dynamic numbers of targets are calculated using tracking algorithm. The proposed people-counting system can, on average, estimate people with an accuracy of 96.11% in the range of 50 to 200 cm. The results were not influenced by changes in lighting condition, except when exposed to sunlight or other strong luminaries according to the experimental results.

References

[1] Masoud, O., Papanikolopoulos, N.P.: A Novel Method for Tracking and Counting Pedestrians in Real-time using a Single Camera. IEEE Transactions on Vehicular Technology 50(5), 1267–1278 (2001)
[2] Chen, T., Chen, T., Chen, Z.: An Intelligent People-Flow Counting Method for Passing through a Gate. In: IEEE Conference on Robotics, Automation and Mechatronics, pp. 1–6 (2006)
[3] Conte, D., Foggia, P., Percannella, G., Tufano, F., Vento, M.: A Method for Counting People in Crowded Scenes. In: IEEE International Conference on Advanced Video and Signal Based Surveillance, Washington, USA, pp. 225–232 (2010)
[4] Chen, F., Zhang, E.: A Fast and Robust People Counting Method in Video Surveillance. In: IEEE International Conference on Computational Intelligence and Security, pp. 339–343 (2007)
[5] Chen, C.H., Chang, Y.C., Chen, T.Y., Wang, D.J.: People Counting System for Getting In/Out of a Bus Based on Video Processing. In: IEEE Conference on Intelligent Systems Design, vol. 3, pp. 565–569 (2008)
[6] Antic, B., Letic, D., Culibrk, D., Crnojevic, V.: K-means Based Segmentation for Real-time Zenithal People Counting. In: IEEE International Conference on Pattern Recognition (2009)
[7] Lu, X., Chen, C.-C., Aggarwal, J.K.: Human Detection using Depth Information by Kinect. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 15–22 (2011)