A Kinect-Based People-flow Counting System

Ching-Tang Hsieh[a], Hui-Chun Wang[a], Yeh-Kuang Wu[b], Liung-Chun Chang[b] and Tai-Ku Kuo[a]
Department of Electrical Engineering, Tamkang University, Tanshui, Taipei,25137, Taiwan [a]
Institute for Information Industry, Taipei, Taiwan [b]
E-mail: hsieh@ee.tku.edu.tw

Abstract —**This paper proposed a bi-directional people-flow counting system with Kinect. It can also be applied to multi-flow to correspond with the demand for the practical application. Firstly, we set the Kinect above the doorway to capture the situation of pedestrian flow. Then this system detects people in the covering area using the depth image information from Kinect system. And we do the morphological processing like erosion to the object and find the region of interest (ROI) often performed on using a mapping-based detection approach. After these previous steps, this system set a detected line and let people go through it. Therefore, we can get people number of the experimental result. For the multi-flow case, it will cause the occlusion problem, so we could apply the depth information to distinguish the target on occlusion problem. Final, we compare the experimental results with the manual count results and other research. Under normal circumstances, our system provides not only almost 100% for bi-directional counting but also correspond with the demand for real-time.**[1]

Keyword: Kinect,8-Connected Components , People-flow counting, ROI.

I. INTRODUCTION

The technique of people counting is an important issue on visual surveillance. Regardless of crowed tracking, crowed cutting or event analysis, etc. In addition, it can also be applied to business, safety and self-control, etc. Therefore, the advantages of computer vision counting development can not only retrieve information of broader covering area but also analyze characteristics of individual objects for people tracking, and then to a multi-function counting system. However, digitization is convenience to retrieve the data by computer analyzing at any time. Following the user demand send the appropriate massage to correspond with the need of total self-control monitoring management. And this system is an important part of intelligent surveillance technology. It has different method about the counting device we have been known.

The traditional manual techniques not only have high labor costs but also cannot work for counting people at the doorway in long-term jobs. And rotating shaft mechanical counting devices are hard to set up also not convenience to use. The devices are only operated by a person. Using electronic devices such as infrared or laser is easier to set-up than mechanical devices. But the electronic devices can be applied to single counting work on each path, still cannot use on the high-flow situation. Nevertheless, follow the creative computer hardware innovation and invention in computer vision techniques advance with time. People find out many specific functional applications using the technique. So, in recent years, there are some people started studying the research about the computer vision application using camera to evolve the counting device, etc. For example, Chen et. al. [1] proposed that base on analyzing some area and color as well as pass through the door or gate to finish a cost-effective bi-directional people-flow counter. But this device cannot effectively solve the ambient light variable problem; Chen et. al. [2] employed dynamic background subtraction module to determine pedestrian objects from a static scene. Not only identify foreground objects as characters, positions and sizes but also do people counting. Although they had improvement on occlusion, the experimental result accuracy is only above 65% also cannot be real-time.

In this paper, we proposed a bi-directional people-flow counting system by Kinect. In recent years, the depth image can be generated by stereoscopic camera, such as in Chang [3]. The Microsoft released the Microsoft Kinect program on 16th, June in 2011. Also, the depth image has smaller noise and is more stable in Kinect than in stereoscopic camera. Not only that, the Kinect use the infrared camera, so it doesn't affect too much in the ambient light variable problem, like Chen [4]. Therefore, we choose the Kinect to take the film in our experiment. We use the depth image to detect the targets by the Kinect. Then do the morphological processing to the targets to get the region of interest (ROI) often performed on using a mapping-based detection approach. After the previous steps mentioned preceding part, we can obtain the result of this experimental.

II. PROPOSED SYSTEM ARCHITECTURE

In this paper, the system's architecture is divided into two parts. First, we do the morphological processing to the depth image of Kinect. And find the target's ROI. Second, determine and calculate the targets found from ROI whether they are in occlusion or not. Then we can count the number of pedestrian. Fig.1 is the flow chart of the proposed system and Fig.2 is the experimental environment.

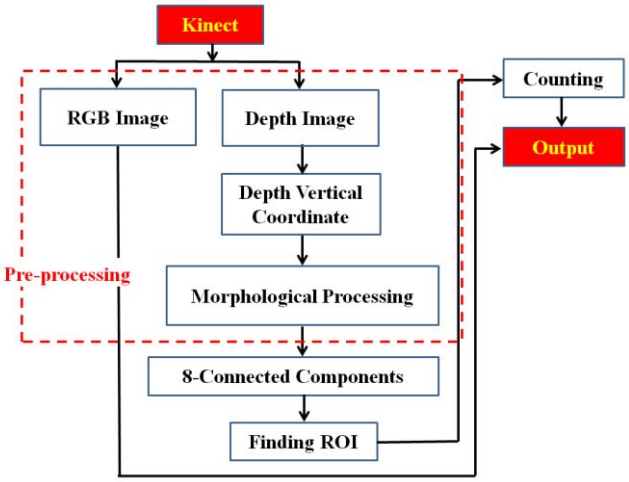


Fig. 1 Flow chart of the proposed system.

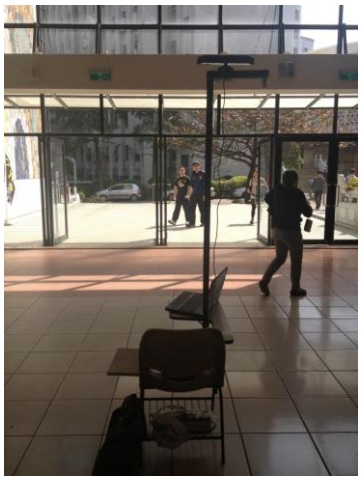


Fig. 2 The experimental environment.

III. PROPOSED METHOD

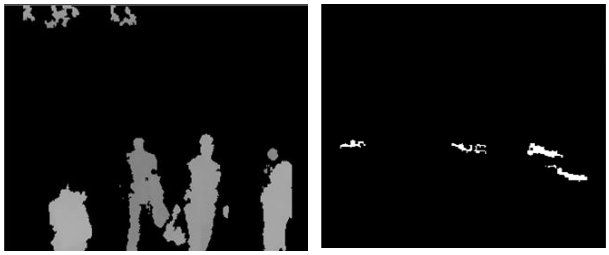
A. Pre-processing

In this system, we can get two kinds of image, one is the RGB image and the other is the depth image as Fig.3. Therefore, we principally use the depth image to do training. However, the depth information is known by us. So we apply the concept of the 3D Cartesian coordinate to make the depth information replace the value of the z-axis. After that, we can get the depth vertical image shown in Fig.4.

The depth vertical image may be contaminated from various types of noise degradations. Hence preprocessing of the depth vertical image is required to remove the background noise. Firstly, we do the morphological processing like erosion on the depth vertical image to denoise. And we do the dilation on the depth vertical image to stretch the targets. Then we can get the targets more clearly. Thus we detected the region of interest (ROI) [5] [6] by using the features as the depth information. We could capture the region of the target and determine it by the threshold whether it is occlusion or not. Fig.5 shows the result after erosion and dilation.

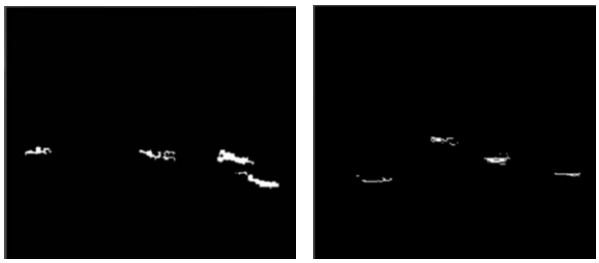
(a) (b)

Fig. 3 (a)The RGB image of Kinect and (b) Depth image.



(a) (b)

Fig. 4 (a)The depth image and (b) The depth vertical image.



(a) (b)

Fig. 5 (a)The depth vertical image before morphological processing and(b) after morphological.

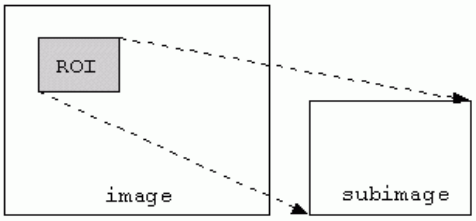


Fig. 6 The illustration of ROI.

B. Region of Interest and Connected-component Labeling

Region of Interest is a rectangular area in an image to segment object for further processing. The illustration is shown in Fig. 6. In the Fig. 6, a ROI is defined at near top left of the image. It is useful to crop an object from an image and to perform template matching within the sub-image, etc. And it is often showed on using a mapping-based detection approach. Note that the ROI has to be inside the image.

In this paper, we use the Connected Component Labeling to find the target's ROI. The Connected Component Labeling can be divided into two types by the definition of connecting. They are 4-Connected Component and 8-Connected Component respectively. In this system, we use the 8-Connected Component.

Connected-component labeling is used in computer vision to detect connected regions in binary digital images, although color images and data with higher-dimensionality can also be processed. When integrated into an image recognition system or human-computer interaction interface, connected component labeling can operate on a variety of information. Blob extraction is generally performed on the resulting binary image from a threshold step. Blobs may be counted, filtered, and tracked. Connected components labeling scans an image and groups its pixels into components based on pixel connectivity, also all pixels in a connected component share similar pixel intensity values and are in some way connected with each other.

Connected component labeling works by scanning an image from top to bottom and left to right in order to identify the connected pixel regions. For example, the regions of adjacent pixels which share the same set of intensity values *V*. For a binary image *V={1}*; however, in a gray-level image *V* will take on a range of values, for example, *V={51, 52, 53, ..., 77, 78, 79, 80}*.

However, for the following we assume binary input images. The connected components labeling operator scans the image by moving along a row until it comes to a point *p* (where *p* denotes the pixel to be labeled at any stage in the scanning process) for which *V={1}*. When this is true, it examines the four neighbors of *p* which have already been encountered in the scan. The neighbors are (i) to the left of *p*, (ii) above it, and (iii and iv) the two upper diagonal terms. Based on this information, the labeling of *p* occurs as follows:

- If all four neighbors are 0, mark a new label to *p*.
- If only one neighbor has *V={1}*, mark its label to *p*.
- If more than one of the neighbors have *V={1}*, mark one of the labels to *p* and make a note of the equivalences.

After completing the scan, the equivalent label pairs are sorted into equivalence classes and a unique label is assigned to each class. As a final step, a second scan is made through the image, during which each label is replaced by the label marked to its equivalence classes. For display, the labels might be different gray-levels or colors. The 8-Connected Component Labeling result shows in Fig. 7and Fig.8.

(a) **(b)**

Fig. 7 (a) The 8-Connected Component Labeling test image and (b) result.

between two people on the occlusion (Fig. 11). Also, it would cause the change of the depth information. Moreover, we could calculate by the depth vertical image to overcome the problem of occlusion. Furthermore, the above cases may be mixed with one another to produce other hybrid cases of

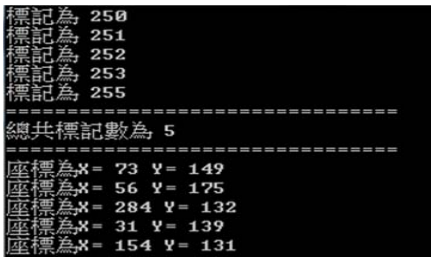
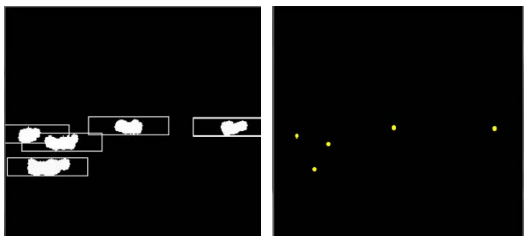


Fig. 8 The 8-Connected Component Labeling result.



(a) **(b)**

Fig. 9 The experimental results of ROI on (a) the depth vertical image and (b) the 8-Connected Component Labeling image.

C. Occlusion Computing

Generally, people may walk alone or together with their friends and this will result in a single-people pattern, several single-people patterns, a multi-people pattern, several multi-people patterns, or their mixed patterns in a surveillance area captured. Therefore, each moving object may happen to be split or merged within a range of the captured surveillance area also this merge-split problem may confuse the results of people counting. Basically, the merge-split circumstances can be classified into four cases, as shown in Fig. 10. A 2-to-1 merging case of Fig. 10 (a) means that two moving targets are merged into a single moving target, where each moving target may be composed of one or more persons. By the same inference, 1-to-2 splitting case of Fig. 10 (b), 1-to-2-to-1 split-merge case of Fig. 10 (c), and 2-to-1-to-2 merge-split case of Fig. 10 (d) mean that one moving target is split into two partial ones, one moving target is firstly split into two partial ones and finally merged into a large one, and two small moving targets are firstly merged into a large one and finally split into two small ones, respectively.

From the above analysis for merge-split, to check if there is a merging or splitting case to happen, the area change of the moving target can give a fundamental judgment. A merging case will be detected if two separate moving targets move in the current frame and then are combined into a larger-area target in the next frame. Thus, the people number of the new merged target needs to be calculated according to those area-based decision rules mentioned above in order to refine the initial count. Based on the opposite reasoning, a judgment of splitting case can be also deduced. But according to the depth vertical image, the occlusion problem could be solved. Therefore, we found that it has strong edge merge-split, in which some cases are intractable to handle. It should be pointed out that the situation of splitting and merging of more than two blobs can be done with the single blob described above.

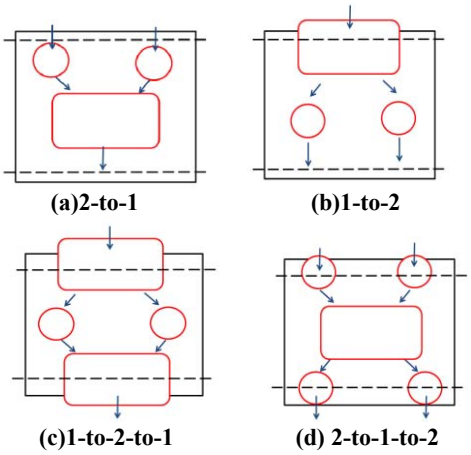


Fig. 10 The illustration of occlusion.

(a) (b)

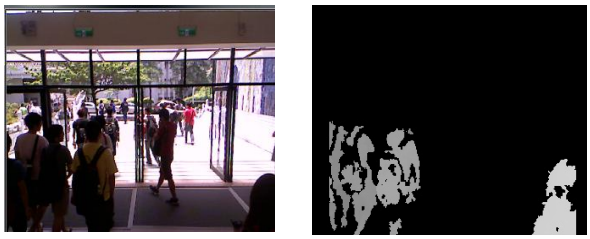
Fig. 11 It has strong edge between two people on the occlusion(a) RGB image (b) Depth image.

A. People Counting

According to the pre-processing steps we mentioned, we can capture the targets and then do the people counting. However, on the people counting, this paper consulted Yam et.al. [8].They proposed a method about setting a detected line in the screen. Let all objects could pass by the line. It not only does calculating but also deposits the direction of the target. We use the depth information to be the detected line. Because the target is closer to the camera, the luminance is lighter. Otherwise, it's darker. Thus, we could calculate the number of pedestrian and the direction of them to get the experimental result and correspond with the effect for real-time.

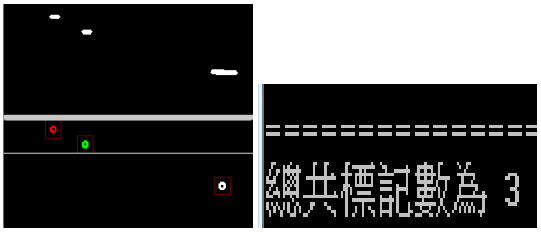
IV. EXPERIMENTAL RESULTS

The hardware used in our experiment is Microsoft Kinect and the computer CPU is Intel® Core(TM) i3-370M 2.4GHz. RAM is 2.4GB. The software we used is Microsoft Visual Studio 2010 and openCV2.3. The input color image and depth image size are respectively 640x480 and 320x240 pixels. Fig.12 and Fig.13 show the experimental results of people counting. Fig. 14 and Fig. 15 show the experimental results of bi-directional people counting.



(a) (b)

Fig.12(a) The result of RGB image and (b) depth image.

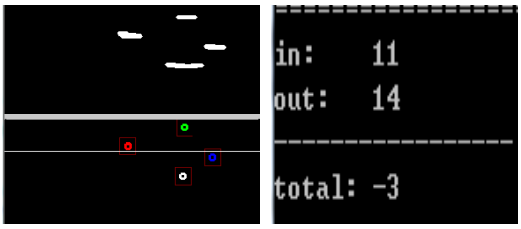


(a) (b)

Fig.13 The experimental result of people counting(a)and(b).

(a) (b)

Fig.14 (a) The result of bi-directional in RGB image and (b) depth image.



(a) (b)

Fig.15 The result of bi-directional people counting(a)and(b).

V. CONCLUSION

This paper used the Kinect set above the doorway to capture the video we needed and the situation of the pedestrian flow. In the end, we compared the experimental result with the manual count result and other research. Under normal circumstance, our system provides not only almost 100% but also correspond with the demand for real-time.

REFERENCES

[1] Thou-Ho(Chao-Ho) Chen, Tsong-Yi Chen and Zhi-Xian Chen, "An Intelligent People-Flow Counting Method for Passing Through a Gate," Robotic, Automation and Mechatronics, 2006, IEEE , pp.1-6.

[2] Chin-Chang Chen, Hsing-Hao Lin and Oscal T.-C.Chen, "Tracking and Counting People in Visual Surveillance Systems," Acoustics, Speech and Signal Processing(ICASSP), 2011, IEEE, pp.1425-1428.

[3] Liang – Chun Chang, "Palm Motion Detection System Based on Stereo Vision," Master paper on Department of Electrical Engineering at Tamkang University, 2010, files approve in 5 years, unpublished.

[4] Guan-Ting Chen, "Hand Tracking System Based on Kinect," Master paper on Department of Electrical Engineering at Tamkang University, 2011, unpublished.

[5] D.Ryan, S. Denman, C. Fookes,"Crowed Counting using Multiple Local Features," Digital Image Computing Techniques and Applications, Dec 2009. DICTA'09, Melbourne, VIC, pp.81-88.

[6] A.G Vicente, I.B. Munoz, P.J. Molina and J.L.L. Galilea, "Embedded Vision Modules for Tracking and Counting People," vol 58,issue 9, IEEE Trans. On Instrumentation & Measurement, Sept, 2009, pp.3004-3011.

[7] R. Fisher, S. Perkins, A. Walker and E. Wolfart, "Connected Component Labeling," website: *http://homepages.inf.ed.ac.uk/rbf/HIPR2/label.htm*

[8] Kin-Yi Yam, Wan-Chi Siu, Ngai-Fong Law, Chok-Ki Chan, "Effective bi-directional people flow counting for real time surveillance system," *in Proc. IEEE* International Conference on Consumer Electronics(ICCE), 2011, pp863-864.