International Conference on Science and Technology 2015, RMUTT

Robust People Counting Using a Region-Based Approach for a Monocular Vision System

Pongsakon Bamrungthai and Suteera Puengsawad
Division of Computer Science
Huachiew Chalermprakiet University
Samutprakarn, Thailand
Pongsakon.Bam@gmail.com, aj-tarn@hotmail.com

***Abstract*—This paper presents an automatic people counting system for indoor environment. The system processes visual information obtained from an IP camera that is mounted in top-view position above the door. The people detection process is designed to perform as a sequence of steps, including background subtraction, morphological operation, and blob detection. The results from the detection process are filtered and tracked to improve robustness of the system. A region-based approach is applied to the people counting process. The image obtained from the camera is divided into three regions and the counting algorithm is developed to track the people's state in the regions. The algorithm can count people with IN and OUT direction. Even when the system was setup without lighting control, the developed system can correctly count people as shown in the experiments. The output of the system can be exported to be used in downstream applications such as security surveillance and statistical analysis of customer behavior.**

***Keywords*—people counting, region-based approach, monocular vision system**

I. INTRODUCTION

In recent years, people information in an environment is increasingly important because of its potential to improve intelligence of the systems. This information includes people presence, count, location, track, and identity. There are many approaches to deal with it as shown in the comprehensive survey of human-sensing [1]. People detection and tracking techniques are applied to extract people information from irrelevant background. One approach is to use monocular vision with perspective view [2,3]. It is rather complex because of random trajectories and occlusions. The multisensor system [4] can provide a very robust performance, but it requires computationally intensive hardware to fuse information from multi-sensors. For indoor environment in [5], multiple depth sensors are used to resolve occlusion, illumination change, and clustered environment problems. The hemiellipsoid head model and the affiliated geodesic human shape are proposed for people detection. The virtual gate and the counting area concepts are proposed in [6] to obtain reliable counts from a single view. The crowd segmentation algorithm is developed to extract individual person from groups of people. The results from the segmentation are used to detect a dangerous event on a subway platform.

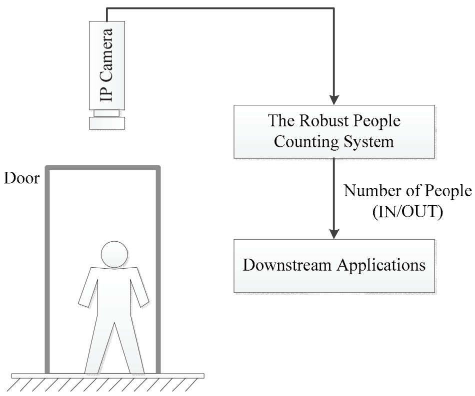

Figure 1. System configuration.

The main objective of this paper is to develop an automatic people counting system for indoor environment. The configuration of the system is shown in Fig. 1. It is a monocular vision system that uses an IP camera mounted in top-view position to provide video stream of people that are walking through a door. The robust people counting system then processes image data to extract people position for tracking and counting both IN and OUT direction. The system is easy to setup. It does not require any other specific hardware device or additional lighting control. The output of the system can be used to estimate people flow information. Many practical applications can be developed based on this work such as security surveillance and statistical analysis of customer behavior.

II. SYSTEM ARCHITECTURE

A. Overview

The system is composed of 3 parts: Image Capturing, People Detection, and People Counting as shown in Fig. 2. The Image Capturing is used to manage video source input that comes from the two sources: 1) live IP camera streaming, and 2) a recorded video file.

978-1-4673-7670-9/15/$31.00 ©2015 IEEE

The image data are transferred to the People Detection to detect and track individual person. The output of the People Detection is the centroid position of the detected person in pixel coordinates. The People Counting is then performed by using a region-based approach as described in section IV. The final output of the system is the number of people with IN and OUT direction.

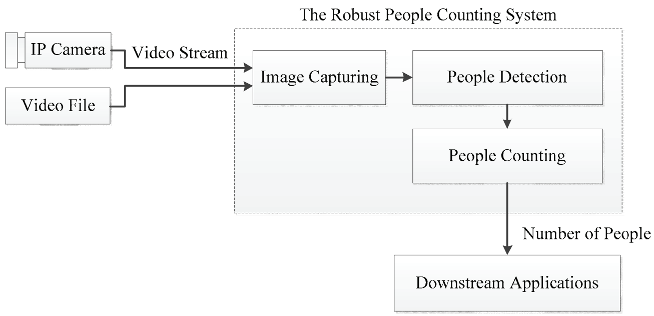

Figure 2. System architecture.

B. The Image Capturing

The details of the two input sources of the Image Capturing are as follows:

- The IP camera used in this paper employs the RTP/RTSP protocol to stream video data to the computer.
- The video streaming that is recorded as a video file can be used as an input to the system. The AVI file format is the standard file format used in this research.

The video data are continually grabbed and decoded from the selected source to get the raw image data for processing.

III. THE PEOPLE DETECTION ALGORITHM

To detect individual person for counting, a people detection algorithm is developed. It is designed as a three-step process that consists of background subtraction, morphological operation, and blob detection as shown in Fig. 3. The details of the algorithm can be described as follows.

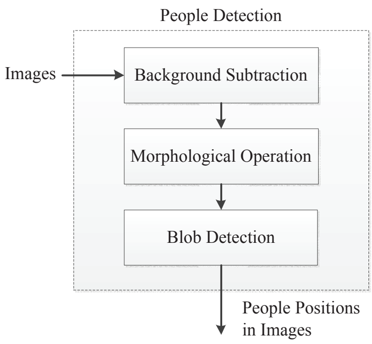

Figure 3. The people detection process.

A. Background Subtraction

A common method to identify moving objects from video stream is to perform background subtraction [7]. A simple approach is to calculate the median intensity at each pixel of the last K frames to create/update the background model. Then the next process is to subtract it from the current image, threshold the difference, and perform some morphological operations to obtain the estimation of moving objects in the scene. But this method is very limited in real applications.

The adaptive Gaussian Mixture Model (GMM) [8] that is a statistical background model is applied in this paper to make the system more robustness to illumination changes and dynamic backgrounds. From the results of the experiments as shown in section V, it is outperformed in people detection when compared with the algorithm of [9]. The result that is a binary image is filtered by using morphological operation as shown in the next section.

B. Morphological Operation

The morphological operation is the approach based on set theory in the field of image processing [10]. Its goal is to segment images into meaningful units. In this paper, it is applied after background subtraction to filter and group related pixel to form objects in the image. Four morphological operations, including erosion, dilation, closing, and opening are tested to get the best result for the counting process. Some results from the testing are shown in section V.

C. Blob Detection

After the morphological operation is applied, the next step is to detect objects that are regions with specified size. The algorithm with linear time searching [11] is used in this paper as the blob detection method. The detected objects with larger or smaller than the specified sizes are then filtered out. The final results are used for the counting process.

IV. THE PEOPLE COUNTING ALGORITHM

To improve robustness of the system, an algorithm has been developed to count people that are walking through a specified region. A region-based approach is applied for this purpose. The details are shown in the following sections.

A. The Counting Regions

The image obtained from the IP camera is separated into 3 regions in the pixel coordinates as shown in Fig. 4. In this paper, the counting is performed in the y-direction. The IN direction of the counting is in the positive y-direction and the opposite is true for the OUT direction.

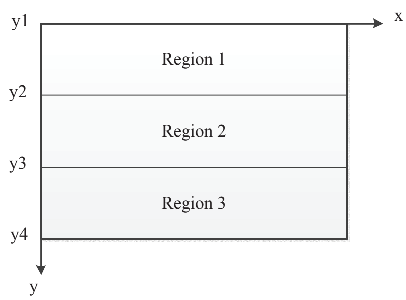

Figure 4. The counting regions in the pixel coordinates.

B. The Counting Algorithm

The people counting algorithm that is developed in this paper is shown in Fig. 5. It estimates the moving direction of the detected people from the states of people in the counting region.

```
Procedure CountPeople(objPos)
if stateIn = 0 AND stateOut = 0 then
    if objPos is in region1 then
        stateIn = 1
    else if objPos is in region3 then
        stateOut = 1
    endif
endif
if stateIn = 1 AND stateOut = 0 then
    if objPos is in region2 then
        stateIn = 2
    endif
endif
if stateIn = 0 AND stateOut = 1 then
    if objPos is in region2 then
        stateOut = 2
    endif
endif
if stateIn = 2 AND stateOut = 0 then
    countIn = countIn + 1
    stateIn = 0
    stateOut = 0
endif
if stateIn = 0 AND stateOut = 2 then
    countOut = countOut + 1
    stateIn = 0
    stateOut = 0
endif
```

Figure 5. The people counting algorithm.

V. EXPERIMENTAL RESULTS

The experiments are performed to find the appropriate detection algorithm for the system. The system captures single person at a time to easily identify the good results. Multi-person counting can be developed further in the future work. Then, the counting algorithm is tested to validate the developed counting algorithm.

A. People Detection

1) Background Subtraction: To find the most effective algorithm for background subtraction, the experiments are performed to compare the two selected algorithms. They are the GMM-based algorithms in [8] and [9]. The results are shown in Fig. 6-7. It is clearly shown that the algorithm in [8] provide better results than the algorithm in [9] because of its clean foreground object extraction.

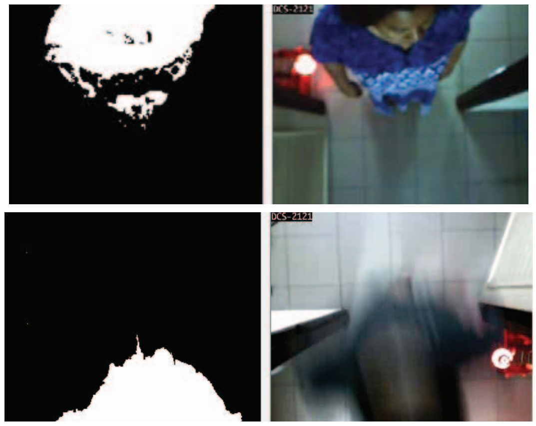

Figure 6. The results of background subtraction using the algorithm in [8].

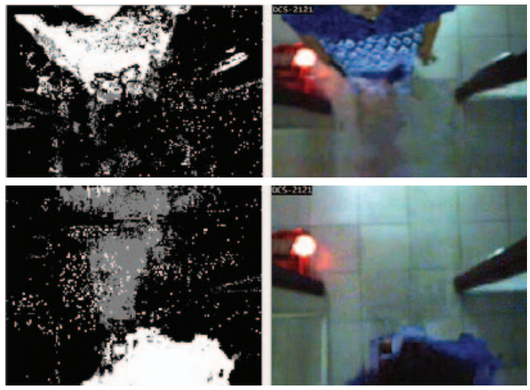

Figure 7. The results of background subtraction using the algorithm in [9].

2) Morphological Operation: The foreground objects in the last step need some operations to filter and segment the pixel data. Various morphological operations are tested to find the appropriate operation. The morphological operations, including erosion, dilation, closing, and opening are tested. The results of closing and opening are shown in Fig. 8-9, respectively. It shows the opening results that are smaller and more suitable for detection and tracking in later process.

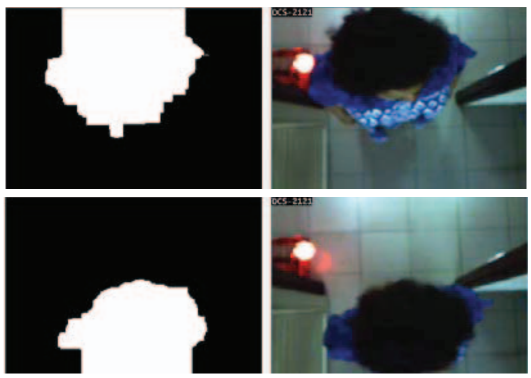

Figure 8. The results of applying morphological closing.

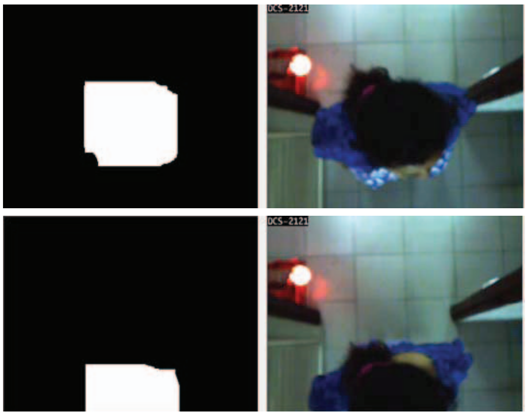

Figure 9. The results of applying morphological opening.

3) Blob Detection: Blob detection is applied after the morphological operation to detect connected component, a group of pixel that belongs to individual objects in image. The results are the centroid positions and areas of the objects in the pixel coordinates. Only the detected objects with the specified size are used for the counting process.

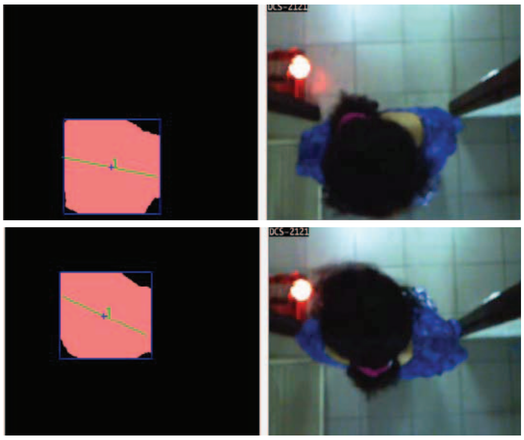

Figure 10. The results of applying blob detection algorithm in [11].

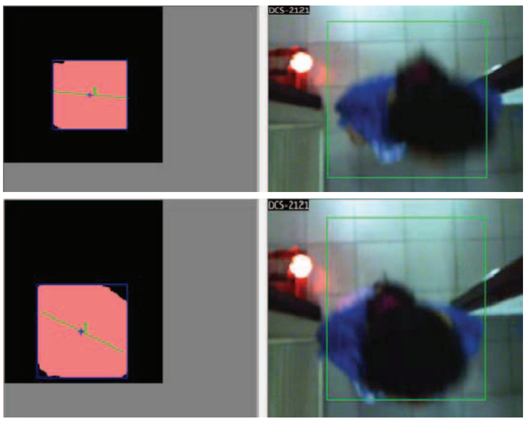

Figure 11. The detected people with ROI during the people count.

Figure 12. The developed system in the counting.

B. People Counting

The people counting is tested by applying the region of interest (ROI) as shown in Fig. 11 to reduce the computational time of image processing. The counting algorithm has been tested by single person walking through a door. The counting test in the developed software is shown in Fig. 12. It is found that the system can detect and count people with IN and OUT direction correctly.

VI. CONCLUSIONS

An automatic people counting system by using monocular vision has been developed in this paper. The algorithms for people detection and counting are designed for real-time processing. The advantage of the system is that it does not require any special processing hardware or lighting control to enable robust counting. By using ROI, the performance of the system has been improved significantly. The experimental results show the appropriate algorithms for people detection and counting. The detection algorithm consists of the GMM-based background subtraction algorithm, morphological opening, and a linear time connected component analysis. It also shows the correctness of the developed system for people counting. The system can be applied to many practical applications after fine tuning the system for performance.

REFERENCES

[1] T. Teixeira, G. Dublon, and A. Savvides, "A Survey of human sensing: Methods for detecting presence, count, location, track and identity," Technical report, ENALAB, 2010.

[2] N. El Nabbout, J.S. Zelek, and D. Clausi, "Automatically detecting and tracking people walking through a transparent door with vision," Canadian Conference on Computer and Robot Vision, pp. 171-178, May 2008.

[3] L. Zini, N. Noceti, and F. Odone, "Precise people countin in real time," 20th IEEE International Conference on Image Processing (ICIP), pp. 3592-3596, Sept. 2013.

[4] D. Schreiber, A. Kriechbaum, and M. Rauter, "A multisensor surveillance system for automated border control (eGate)," 1st Workshop on Activity monitoring by multiple distributed sensing (AMMDS) at the 10th IEEE Int. Conf. on Advanced Video and Signal-based Surveillance (AVSS), 2013.

[5] T. E. Tseng, A. S. Liu, P. H. Hsiao, C. M. Huang, and L. C. Fu, "Real-time people detection and tracking for indoor surveillance using multiple top-view depth cameras," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2014.

[6] X. Liu, P.H. Tu, J. Rittscher, A. Perera, and N. Krahnstoever, "Detecting and counting people in surveillance applications," in Proc. Advanced Video and Signal based Surveillance (AVSS), pp. 306-311, 2005.

[7] M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis, and Machine Vision, 4th ed., Cengage Learning, 2015.

[8] P. KaewTraKulPong and R. Bowden, “An improved adaptive background mixture model for real-time tracking with shadow detection,” in Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, Sept. 2001.

[9] Z. Zivkovic, “Improved adaptive Gaussian mixture model for background subtraction,” 17th International Conference on Pattern Recognition, vol. 2, pp. 28-31, 2004.

[10] R.C. Gonzalez and R.E. Woods, Digital Image Processing, 3rd ed., Upper Saddle River, NJ: Pearson Prentice Hall, 2008.

[11] F. Chang, C.-J. Chen, and C.-J. Lu, “A linear-time component-labeling algorithm using contour tracing technique,” Comput. Vis. Image Understand., vol. 93, no. 2, pp. 206-220, Feb. 2004.