Single-pixel approach for fast people counting and direction estimation

A.O. Adegboye, G. P. Hancke and G. P. Hancke Jr.[1],
Department of Electrical, Electronic and Computer Engineering,
University of Pretoria, Private bag X20 Hatfield, Pretoria 0028
Tel: +27 12 420 3736, Fax: +27 12 362 5000
email: {dolly.adegboye, gerhard.hancke}@up.ac.za; ghancke@googlemail.com[1]

***Abstract*- In recent years, people detection and counting has attracted a lot of attention in visual surveillance and security. However, it remains a challenging and complex task for cases where occlusions, varying illumination and weather conditions are experienced. In order to bypass and resolve these challenges, a two-part method based on feature extraction is proposed. In the first part, a single-pixel method for background segmentation is proposed, and in the second part, a virtual-line direction-estimation method is proposed where the direction in which the person is moving is estimated before counting. This method aims to overcome the shortcomings of people detecting, tracking and counting methods.**

***Index Terms*- people counting, direction estimation, multi-target tracking.**

I. INTRODUCTION

Nowadays, the need for autonomous systems is becoming more and more important because there is a rising demand in security applications where video surveillance plays a critical role. Manually-operated systems are becoming less applicable because these systems are usually affected by human-related factors such as fatigue and stress. As a result, extensive studies are continuously performed in order to highlight the problems faced and provide solutions for the surveillance systems.

People counting algorithm is one of the crucial and challenging problems in video surveillance because it is strongly dependent on other image processing algorithms. It is an estimation process where the number of people present in a scene are counted and recorded. It is mostly implemented for security reasons such as: 1) monitoring the number of people that enter or exit a building so that when an emergency occurs, it will be easier to know how many people that are in the building and possibly where to look; or 2) for traffic surveillance where pedestrian traffic flow can be estimated [1].

Typical scenes vary from outdoor to indoor areas such as shopping malls, buildings and underground parking areas, streets, private residents, hotels. Fig. 1(a) - (c) provide examples of the typical scenes. Although people counting algorithm has attracted a lot of attention, it remains a challenging and complex task for cases where occlusions, varying illumination and weather conditions are experienced.

(a) (b) (c)

Fig. 1. Typical people counting scenes (a) Wembley Stadium. (b) Airport entrance/exit. (c) Shopping mall entrance/exit.

II. RELATED WORK

There are two methods of implementing people counting algorithms, namely detection-based and feature extraction-based methods [1], [2]. Detection-based methods involve the process of individually detecting all the people present in the scene where the entire person or parts of their bodies can be detected [1], [3] or each person can be reconstructed using human 3D shapes [4]. Once detection has occurred, the total number of detected persons is counted and recorded.

Feature extraction-based methods, also known as map or measurement-based methods [1] involve processes where various features are extracted from the video frames and the useful features are grouped according to their similarities into a feature vector. Then, the feature vector is employed to count the number of persons in the scene. Feature extraction-based methods can be further subdivided into: motion detection and analysis-based methods [5], template-based methods [6], spatio-temporal methods [7], feature trajectories clustering [8] and low-level feature regression methods [1], [3], [9] - [11].

Feature extraction-based methods are the preferred approaches because they are more robust, preserve privacy and not primarily dependent on detection. However, people-counting algorithms face challenges which include partial occlusions, imperfect foreground extraction, false detections, illumination changes, human pose and direction. Several attempts have been made to either solve [8], [12], or bypass the challenges using hardware-based methods and direction-based methods [11]. Most of these attempts are application-specific so they fall short in robust cases. Also, there is a trade-off between the complexity and performance accuracy of the algorithm, that is, the higher the performance accuracy the higher the complexity becomes. As a result, the proposed approach focuses on implementing a fast and robust people-counting algorithm with low complexity and high performance accuracy which is invariant to occlusions and illumination conditions.

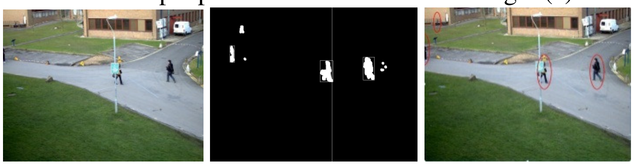
III. METHODOLOGY

The proposed approach is divided into two parts: single-pixel method and virtual-line direction estimation method.

A. *Single-pixel method*

Single-pixel method is further divided into four steps: video conversion, background subtraction, morphological image processing, and single-pixel process. In the first step, the input video is converted into image frames and the initial frame is read as background image as shown in Fig. 2(a). In the second step, the background image is subtracted from each current image in order to produce foreground images. The images are converted into binary based on a threshold value in order to obtain foreground pixels.

Afterwards, the foreground pixels are morphologically processed and represented as bounded blobs as shown in Fig. 2(b). Then, the area values of each blob are calculated and the *xy*-centroid coordinates are extracted from each blob in the input images. These local features are fed into the virtual-line direction estimation method in order to estimate the number of people in a scene as shown in Fig. 2(c).



(a) (b) (c)

Fig. 2. PETS Dataset (a) Input image. (b) Bounded blob binarized image. (c) Result image based on single-pixel method.

B. *Virtual-line direction estimation method*

The implementation of the virtual-line direction estimation method is achieved through two steps: direction-estimation method and distance-computational counting.

In the direction-estimation method, a vertical or horizontal virtual line is drawn at the entrance point of the scene. Then, the *xy*-coordinates of the virtual line are subtracted from the *xy*-coordinates obtained from the single-pixel method. If output value is less than 0, it is assumed that the individual is heading towards the entrance point. So the direction is estimated towards the entrance point of the scene. Otherwise, it is assumed that the person is heading away from the entrance point of the scene.

Distance-computational counting is the process where the total distance between the blob and virtual line is calculated so that the blobs can be categorized as either short-distanced or long-distanced. Then, the number of people that is contained in each blob is calculated according to their distances.

IV. PERFORMANCE EVALUATION

The proposed approach is evaluated by using mean square error (MSE) computation. This is the process where global difference is computed between input and output values, and then an error percentage of the difference is returned. For input values, the total number of people in each scene is manually counted and recorded as the ground truth, while total number of people obtained from the proposed is regarded as the estimated truth for the output values. MSE is defined in eq. 1 as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}[\hat{C}(t) - C(t)]^2 \qquad (1)$$

where N is the total number of scenes, C is the ground truth and $\hat{C}$ is the estimated truth.

V. SUMMARY AND CONCLUSIONS

Currently, the *xy*-centroid coordinates and area value of each blob are recorded and saved into text files and these features are used to estimate the directions of the people. Also, the single-pixel method is implemented using a simple background segmentation approach. In the future, multi-target particle filter will be implemented as an addition to the single-pixel method in order to monitor and record the location of each person.

VI. REFERENCES

[1] X. Wang, C. Wang and J. Yao, "A Heuristic Information Based System for People Counting," *in 2011 International Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 1, no. 1, pp. 22-26, Aug. 2011.

[2] Y. Benabbas, N. Ihaddadene, T. Yahiaoui, T. Urruty and C. Djeraba, "Spatio-Temporal Optical Flow Analysis for People Counting," *in 2010 IEEE International Conference on Advanced Video and Signal Based Surveillance*, vol. 1, no. 1, pp. 212-217, Aug. 2010.

[3] H. Xu, P. Lv and L. Meng, "A people counting system based on head-shoulder detection and tracking in surveillance video," *in 2010 International Conference on Computer Design and Applications*, vol. 1, no. 1, pp. 394-398, Jun. 2010.

[4] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1198-1211, Jul. 2008.

[5] X. Liu, P. H. Tu, J. Rittscher, A. Perera, and N. Krahnstoever, "Detecting and counting people in surveillance applications," *In International Conference on Advanced Video and Signal-Based Surveillance*, vol. 1, no. 1, pp. 306-311, Sept. 2005.

[6] O. Sidla, Y. Lypetskyy, N. Brandle, and S. Seer, "Pedestrian detection and tracking for counting applications in crowded situations," *In International Conference on Advanced Video and Signal-Based Surveillance*, vol. 1, no. 1, pp. 70-75, Nov. 2006.

[7] A. Albiol, I. Mora, and V. Naranjo, "Real-time high density people counter using morphological tools," *IEEE Transactions on Intelligent Transportation Systems*, vol. 2, no. 4, pp. 204-218, Dec. 2001.

[8] Ya-Li Hou, Grantham K. H. Pang, "People Counting and Human Detection in a Challenging Situation," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 1, pp. 24-33, Jan. 2011.

[9] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, "Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking," *In IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 1, pp. 1-7, Jun. 2008.

[10] P. Kilambi, E. Ribnick, A. Joshi, O. Masoud, and N. Papanikolopoulos, "Estimating Pedestrian Counts in Groups", *Computer Vision and Image Understanding*, vol. 110, no. 1, pp. 43-59, Apr. 2008.

[11] J. Xing, H. Ai, L. Liu and S. Lao, "Robust crowd counting using direction flow", *on 2011 8th International Conference on Image Processing*, vol. 1, no. 1, pp. 2061-2064, Sept. 2011.

[12] A. G. Vincente, I. B. Munoz, P. J. Molina and J. L. L. Galilea, "Embedded vision modules for tracking and counting people", *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 9, pp. 3004-3011, Sept. 2009.

Adedolapo Adegboye received her B. Eng. degree in computer engineering from the University of Pretoria in 2010. She is presently studying towards her Master degree in Computer Engineering at the same institution. Her current research interests include visual surveillance, image processing, pattern recognition and computer vision.