Effective bi-directional people flow counting for real time surveillance system

Kin-Yi Yam[1,2], Wan-Chi Siu[1], Ngai-Fong Law[1], Chok-Ki Chan[2]

[1]Department of Electronic & Information Engineering, The Hong Kong Polytechnic University, Hong Kong

[2]TeleEye Holding Limited (HK), Hong Kong

Abstract — ***In this paper, an object based bi-directional counting system is proposed, which comprises of an advanced object detection and tracking algorithm to count the people flow in the monitoring scene. The result of which shows that the approach can provide about 90% accuracy for bi-directional people counting with an angle of 45^o to the scene.***

Index Terms — ***Surveillance system, counting algorithm, bi-directional counting, scene calibration and object detection.***

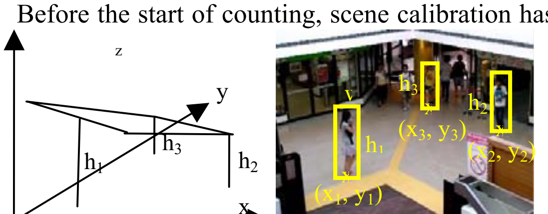
I. INTRODUCTION, REVIEW AND OVERVIEW

People flow counting is one of the important additional information items for surveillance systems. Many researchers have made efforts on the topic. It can be classified into two main approaches. (1) Line Of Interest (LOI) detection - an overhead camera with LOI is the most commonly used setting in order to reduce the occlusion problem in recent works[1,2]. However, it is hard to install an ideal overhead camera in low ceiling building. (2) Region Of Interest (ROI) detection - the counting is often performed on using a mapping-based detection approach[3,4]. However, these kinds of counting methods can only estimate the number of people in a region, which is lack of directional information of people flow.

In order to solve this problem, we propose in this paper an efficient and easy to implement counting approach for real time monitoring systems. This can be used as a plug-in module to an existing surveillance system that is usually installed with an angle of 45o to the monitoring scene. Our system consists of three parts: (i) an N-bins histogram double background modeling and subtraction algorithm, (ii) a camera calibration and object detection method, and (iii) a bi-directional flow counting of objects using a detection line. The background modeling and object detection parts are based on our previous work[5].

II. SCENE CALIBRATION

In our counting system, a normal surveillance angle of 45^o is used. Before the start of counting, scene calibration has to be



(a) (b)

Fig.1 (a)The 3D head plane model, (b)Scene calibration by three rectangles

applied, as shown in fig.1. Three points have to be selected in order to generate a linear 3D human head plane on the monitoring scene, which is shown in fig.1(b). This plane represents the height of a person at different positions on the screen. The values of x and y can be found by the object detection method used, so that the only unknown, z, can be obtained. Also the human width-height ratio can also be approximated for reference. This ratio is used to estimate the width of the detected object according to the height interpolated from the 3D head plane.

III. OBJECTS DETECTION METHOD.

Our object detection is divided into two parts: IOD and EOD.

A. *Initial Objects Detector (IOD):* The object detection method is based on the error frame obtained by the background subtraction approach in our system[5]. A matrix detector is used to detect the potential size of an object. Once an object is found, a boundary box is generated around the object with the boundary box location (x, y) coordinate values. These values are substituted in the liner 3D head plane equation to obtained z which represents the height of the object at that position. Compared with the possible height and the z value of a person recorded before, if the height is equal to or within the tolerance, a person will be identified, and at the same time a new object is created. After the person has been detected and initialized, appropriate parameters are stored as the object features, such as the object center, size, initial position and the color histogram. All these features are used for object identification and tracking.

B. *Existing Object Detector (EOD)*: All new objects found from the IOD are stored in a memory buffer. Every time when an error frame is entered into the detection module, the EOD is implemented firstly. The search starts from the existed object center, and moves spiral outward to the detected object. When a pattern is detected, the system applies the same process as IOD. But the size of the detected object will be compared with the predicted size of an object according to the position. If the size is equal to or within the tolerance, the detected object is confirmed the same as the existing object. If not, a group of people is detected, which is to be discussed in the next section. Once the same object is detected again, the new object center, size and color histogram feature are extracted. Then the object features are updated according to the following equations.

$$O(t)_{centre} = O(t-1)_{centre} + (O(t-1)_{centre} - O(t)_{centre}) \times \alpha \quad (1)$$

$$O(t)_{width} = O(t-1)_{width} + (O(t-1)_{width} - O(t)_{width}) \times \beta \quad (2)$$

$$O(t)_{height} = O(t-1)_{height} + (O(t-1)_{height} - O(t)_{height}) \times \beta \quad (3)$$

where O is the detected object, t is the current time, α and β are updating factors <1 and set to 0.8 and 0.75 respectively.

C. *Detect groups of people*: Fig.2 shows that there are three persons merging together in the middle of the scene. As we have the prior knowledge of human size at that position, this error pattern can be cut into a number of persons according to the object within the large pattern as shown.

Fig. 2 Groups of people detection

Then the “prelim objects” can be detected. The term “Prelim object” in the scene means that the detected object has not yet been verified. In order to look for the same object in the pattern, feature matching with histogram [6] and centre position is used.

D. *Object Counting*: Our counting definition is that when a person passes through the detection line, the counter will be incremented by one and it also records the direction of movement. The detection line can be written in intercept form, y = mx +c, where m is the slope and c is the y-intercept. For every object, we can calculate the perpendicular distance between the detection line and the mid point of human feet (p, q). Let us make use of the 2D Point Line distance [7] as shown below. Rewrite the general equation of the detection line as:

$$aX + bY + c = 0, \quad or \quad Y = -\frac{a}{b}X - \frac{c}{b} \tag{4}$$

where a, b and c are the constants. Hence, points (x, y) on the detection line form a vector:

$$\begin{bmatrix} x \\ -\frac{a}{b}x - \frac{c}{b} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{c}{b} \end{bmatrix} - \frac{1}{b}\begin{bmatrix} -b \\ a \end{bmatrix} x \tag{5}$$

Thus, the vector $\vec{m} = [-b \quad a]^T$ is parallel to the detection line. Consider the point (p,q) to the detection line. Vector $\vec{v} = [a \quad b]^T$ is perpendicular to m. Then we can express the vector from the point to the line as:

$$\vec{r} = \begin{bmatrix} x - p \\ y - q \end{bmatrix} \tag{6}$$

Projecting r onto v, we have:

$$D = \frac{|\vec{v} \bullet \vec{r}|}{|\vec{v}|} = \frac{|a(x-p) + b(y-q)|}{\sqrt{a^2 + b^2}} = \frac{|ap + bq + c|}{\sqrt{a^2 + b^2}} \tag{7}$$

where D is the perpendicular distance between a point to the detection line. If we remove the absolute of the numerator, D becomes a signed value. The sign can be used to represent the direction of the object. Based on this formulation, two kinds of information can be obtained. 1. The distance between each person and the detection line: if the value approaches to zero, it means that the object is close to the detection line. 2. The direction of people passing through the detection line: if the sign is different from the initial sign of the object, it means that the object is passing through the detection line.

Hence, once the sign of the object changes, it means that the object is passing through the line. The following shows the mathematical expression of the counting method.

$$O^0_{sign} = initial\,sign$$

$$O^t_{sign} = \begin{cases} O(t)_{sign} & if[(O^{t-1}_{sign} \neq O(t)_{sign}) \cap (O(t)_{LineDistance} > Threshold)] \\ O^{t-1}_{sign} & othervise \end{cases} \tag{8}$$

$$Count = \begin{cases} Count + 1 & if\ O^t_{sign} \neq O^{t-1}_{sign} \\ Count & othervise \end{cases} \tag{9}$$

where O^t_{sign} and $O(t)_{sign}$ are the sign parameter and the calculated sign respectively of an existing object at time t, and $O(t)_{LineDistance}$ is the distance between an existing object and the detection line. The threshold is set to avoid false counting.

IV. Experimental result

Fig.3 shows three scenes of many monitoring scenes tested. The first two sequences were taken from the university campus with a camera with almost 45° viewing angle. The third sequence represents a real-time surveillance system, placed with an angle of almost 40°. Table 1 shows an accuracy of around 90% in both directions. Note that sequence 3 is a shopping mall. Due to the low ceiling level, the sizes of people are changing sharply. However, our 3D head plane can still estimate the height of human at different positions in the scene. Our system provides a good tracking performance which leads to accurate counting results. The results also show that our proposed method gives better counting for complicated scenes.

Table 1.The Accuracy of the testing sequences.

Accuracy	Sequence 1: Campus1	Sequence 2: Campus 2	Sequence 3: Shopping mall
People flow from bottom to top	89.47%	88.89%	91.43%
People flow from top to bottom	95.56%	96.30%	89.52%

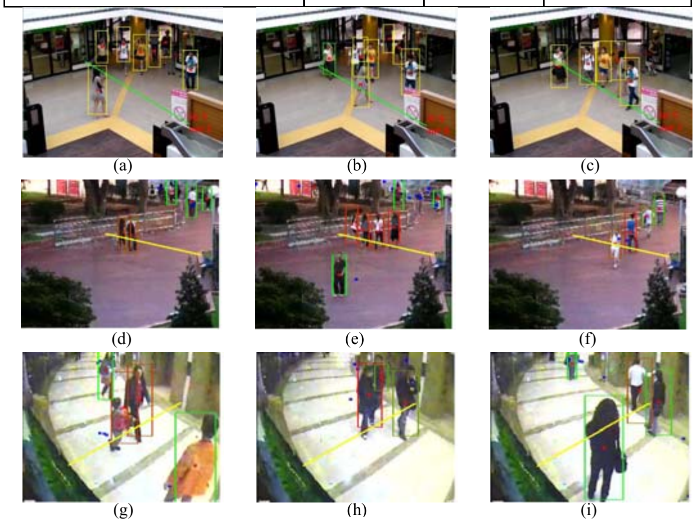


Fig. 3 (a–c) Sample frame from sequence 1. (d-f) Sample frames from sequence 2. (g-I) Sample frames from sequence 3.

V. Conclusion

To conclude, we have proposed a real time and practical counting algorithm for a surveillance camera with normal mounting. Our system can act as a plug-in module to any existing surveillance with the minimal camera setting and provide accurate counting results of directional people flow. This is achieved by combining our feature matching and point line distance techniques on the object and the detection line.

VI. Reference

[1] J. Barandiaran, B. Murguia and F. Boto, “Real-Time People Counting Using Multiple Lines”, Proc., pp. 159-62, 9th Int. Workshop on Image Analysis for Multi. Interactive Services, 7-9 May 2008, Klagenfurt.

[2] S. Yu ; X. Chen ; W. Sun ; D. Xie, “A robust method for detecting and counting people”, pp 1545-9, Int. Conference on Audio, Language and Image Processing, 7-9 July 2008. ICALIP 2008, Shanghai

[3] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, “Crowd Counting using Multiple Local Features”, pp81-8, Digital Image Computing: Techniques and Applications, Dec 2009. DICTA '09, Melbourne, VIC.

[4] A.G. Vicente, I.B. Munoz, P.J. Molina and J.L.L. Galilea, “Embedded Vision Modules for tracking and Counting People”, pp.3004-11, vol 58, issue 9, IEEE Trans. on Instrumentation & Measurement, Sept, 2009.

[5] K.Y. Yam, W.C. Siu, N.F. Law and C.K. Chan, “Fast Video Object Detection Via Multiple Background Modeling”, ICIP, 2010

[6] D. Comaniciu and P. Meer. “Kernel-Based Object Tracking”, pp 564-77 Vol. 25, IEEE Trans. on pattern analysis & Machine Int., May 2003.

[7] C. Kimberling, "Triangle Centers and Central Triangles." Congr. Numer. 129, 1-295, 1998.