

Real-Time People Counting Using Multiple Lines

Javier Barandiaran, Berta Murguia and Fernando Boto
VICOMTech Paseo Mikeletegi, 57
20009, San Sebastián, Spain
{jbarandiaran, bmurguia, fboto}@vicomtech.org

Abstract

A novel real-time people counting system is presented in this paper. Using a single overhead mounted camera, the system counts the number of people going in and out of an observed area. Counting is performed by analyzing an image zone composed by a set of virtual counting lines. The system runs on a commercial PC, does not need a special background and is easily adjustable to different camera height requirements. We have tested the performance of the system, achieving a correct people counting rate of 95%.

1. Introduction

People tracking and counting is a research field that has gained a lot of attention in the last few years. There are a lot of surveillance cameras already installed around us but there are no means to monitor all of them continuously. Therefore it is necessary to develop computer vision based technologies that automatically process those images, in order to detect problematic situations or unusual behaviors.

This work is aimed to solve the problem of counting people that enter and exit a place. This is an important task in order to increase security, to control capacity, or to design marketing strategies.

Traditionally this problem has been solved using turnstiles, laser or other kinds of sensors that are very intrusive and limit people's freedom of movement. As opposed to this, computer vision offers a non-intrusive and cheap solution.

We have developed a real-time system using a CCTV overhead camera that counts people with a correct people counting rate of 95%. The remainder of the paper is structured as follows. In section 2 we review different solutions proposed in the literature. In section 3 our system is described and in section 4 results obtained from several tests are presented.

2. Related work

There have been many approaches trying to solve the problem of people counting. One solution is tracking people while they pass from one region of the scene to another. Most of the works track blobs obtained from a motion detection process [1-3]. Another approach is based on shape detection or recognition [4-6]. These approaches aim at detecting people by searching for heads, legs or silhouettes. Recently a new solution specially suitable for crowded situations has been presented. This solution is based on feature points clustering [7,8] with the objective to identify each moving entity thanks to their independent motion.

Frequently an overhead camera is used to make the counting task easier by avoiding occlusions between people and simplifying person model. For example in [4], the counting is done by means of head detection, but depending on hair and clothes colours, the heads can show a not enough recognizable shape. In [9] a high-density people counting system specifically designed for a train entrance using an overhead camera is presented. This system needs three lines with specific background colour. While people enter, these horizontal lines of every frame are stored onto separate stacks. Once the train doors are closed the stacks are analysed and people counted using morphological tools to separate their blobs. As opposed to this, our system does not need the complete sequence to perform the counting, instead it counts on the fly and does not require any specific ground colour. In [10] two lines are analysed, using DCT based segmentation foreground blobs are extracted, separated using morphological tools and counted as in or out depending on which of the lines is crossed first.

Our system uses multiple independent counting lines, it is not based on the order in which the lines were crossed to know if the person is going in or out, which can be problematic in some circumstances, instead the optical flow is used. Another important

improvement of our system against [9] and [10] is that it does not use morphological operators, because trying to separate a group of people into individuals is not always possible.

3. System description

In this section we describe the proposed solution to the problem of counting people who are crossing a predefined observed area in an image. Input images are obtained from a single camera of either mono or multi-channel type mounted on an overhead position.

The main idea of the solution consists in the definition of an area of interest in the images where motion is being analysed. This area or counting zone is manually defined by the user wherever in the image, by specifying the desired orientation, width (czw) and length. Using a configuration GUI the user must also choose the person's width and distance between lines accordingly. The distance must be greater than half of a person's width. This is done by drawing some lines over a picture when somebody is standing in the middle of the counting zone. Within the area of interest, the configuration process establishes a fixed number of equidistant virtual lines which are placed orthogonally to the expected direction of movement. In Figure 1 a graphical description of the zone is presented showing the parameters involved in its construction. Each parameter is measured in pixels.

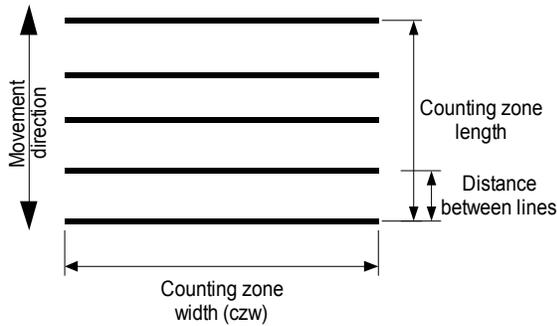


Figure 1. The counting zone

The system is applicable to indoor scenarios, like corridors or entrances where people can go in or out always in the same direction, i.e. in a parallel way to the expected movement direction.

The algorithm is divided in three different steps. Firstly, motion is detected, and moving regions extracted. Then, counting is accomplished by each line. Finally, a global analysis of the results obtained for each line is performed.

3.1. Motion detection

Motion is detected using image differencing between consecutive frames. In order to avoid false positives due to image noise, the difference image is thresholded. Frame differencing is the simplest technique to extract moving information from the background, nevertheless it is very effective for counting tasks and has very low computational cost. Furthermore this method, since it only uses the previous frame, has a faster adaptation against illumination changes than the adaptive background based methods. However, while it works properly with soft shadows, it is easily affected by direct lighting producing hard shadows.

3.2. Counting lines

As mentioned before, counting is performed independently for each line belonging to the counting zone. Each line is represented by the function l , where abscissa and ordinate axis corresponds to the position inside the line and number of accumulated foreground pixels respectively.

$$l_x^t = l_x^{t-1} + D_{i,j}^t \quad \forall x \wedge 0 \leq x < czw$$

$$l_x^0 = 0 \quad \forall x \wedge 0 \leq x < czw$$

In frame t the function value at point x , where x goes from one point of the line to the other, is equal to the previous value in frame $t-1$ plus the value of the corresponding pixel i,j in the difference image D , that can be one or zero. The relation between the points x and pixels i,j is automatically defined when the user draws the counting zone over the image.

$$D_{i,j}^t = \begin{cases} 1 & \text{if } |I_{i,j}^t - I_{i,j}^{t-1}| > \text{threshold} \\ 0 & \text{otherwise} \end{cases}$$

The difference image D^t , as said in the previous section, is the thresholded difference image between the new frame I^t and previous one I^{t-1} . After a short period of time without motion being detected over the line, its function is reset to zero.

When someone crosses a line, pixels are accumulated as in the line function is described. Crossing people are detected by analyzing each line function and detecting intervals of non-zero values. The line counter is incremented when a new interval with enough length is detected, being the length the number of line points occupied by the interval. The length must be bigger than half of a person's width.

For the count to be accurate the algorithm must deal with multiple people, and must determine the direction

of motion for each person. The solutions proposed to cope with these problems are stated below.

In order to know the direction of movement of a person when it is crossing the line, as in [9], optical flow is calculated, only inside a region of interest around the counting zone, using Lucas-Kanade method. The motion vector is estimated calculating the average of the optical flow in a region around the interval and taking into account only the pixels where motion is detected. Finally, the dot product between this direction and the normal movement direction of the zone is computed to know if the person is going in or out.

One particular case arises when two or more people cross one line simultaneously. If the distance separating the persons is too low, the corresponding detected intervals will overlap and the number of persons is then determined from the length of the overall interval and person's width.

Another particular case is when two people cross one after the other without any separation between them. Thus, the function does not become zero once the first person has crossed because the second one occupies the same interval. This problem is very frequent for example in crowded situations or queues. To overcome this problem a reset time is needed to decide when another person must be counted. This time depends on movement velocity. If the time is too small, a slow person could produce multiple counts. Therefore the time is calculated independently for each interval based on the magnitude of the optical flow, the height of the camera and the estimated person's width.

3.3. Combining counting lines

The final step of the algorithm consists in the global analysis of the counting, by combining the results obtained for each line. The count calculated for the whole region of interest is equal to the count supported by the maximum number of lines. In this way a more robust counting zone is obtained through the combination of multiple independent counting lines. When motion is not detected, the count of each line is reset to eliminate possible accumulated errors in some of the lines, for example produced by someone walking orthogonally to the expected direction.

4. Results

Several tests in three different scenarios have been realized in order to measure the performance of the system. The main difference between scenarios is the ceiling height, varying from three to four meters. The higher the camera is placed, the wider the counting

zone is, allowing to test more different and difficult situations. In Figure 2 some frames of the tests in each scenario are shown. Lighting conditions were also different for each scenario, although all of them were interior so differences are not representative.

For every test the same low quality wireless CCTV camera with 3.5mm focal length and a resolution of 352x288 pixels has been used.



Figure 2. Test scenarios 1, 2, and 3 with different heights (3m, 3.5m and 4m respectively)

During the tests we evaluated different situations: one person passing alone with different velocities, multiple people crossing at the same time in opposite directions and groups of people going together. An estimation of the system precision can be done by comparing the real count against the system count. However, it is not suitable in our case because false positives compensate false negatives. Instead, we have analysed the confusion matrix where tp (true positives) are the successes of the system. The errors are fn (false negatives) and fp (false positives), where fn represents someone that was not counted and fp someone that was counted twice. The confusion matrix allows to estimate the precision and recall as follows:

$$precision = \frac{tp}{tp + fp} \quad recall = \frac{tp}{tp + fn}$$

These are standard measures for evaluating the quality of a classification process. While *precision* measures the exactness of the process, *recall* measures the completeness. The measure F , called the weighted harmonic mean, is a way to combine *precision* and *recall* obtaining a general quality measure.

$$F = \frac{2 * precision * recall}{precision + recall}$$

The obtained results are shown in Table 1. The column called *real* represents the real number of people that went in and out against the count performed by the system in the next column. The third and fourth columns show the number of correct detections (tp) and incorrect detections (fn and fp). The last columns contain the resultant *precision* (p), *recall* (r) and F measures. As can be seen, the mean F is equal to 0.95 with a total of nearly one thousand people.

It is important to notice that we have not only tested simple situations, but also some complex and difficult

enough ones, for example groups of people crossing continuously at the same time in opposite directions, in order to obtain a realistic error estimation. However it would be necessary to perform more diverse tests with different camera heights and lighting conditions. An important conclusion extracted from the tests is that the majority of errors are false negatives, *recall* is lower than *precision*, mainly because of errors in the motion detection process and people crossing near to the counting lines edges. Furthermore an interesting result is shown in T3 where the system seems to be more reliable but with lower *recall*.

Table 1. Tests results

	<i>real in+out</i>	<i>system in+out</i>	<i>tp</i>	<i>fn+fp</i>	<i>p</i>	<i>r</i>	<i>F</i>
T1	101+95	97+90	182	14+5	0.97	0.93	0.95
T2	232+241	225+233	445	28+13	0.97	0.94	0.96
T3	127+128	116+117	231	24+2	0.99	0.91	0.95

Difficult situations for our system can be represented by children or people carrying things, as it only relies on the detected intervals length to decide the number of counted people. Also the system can have problems with people moving very slowly, mainly because the motion detection method would not segment enough foreground pixels. If someone crosses the zone horizontally it would produce errors, depending on the number of crossed lines.

Some results videos can be downloaded from http://www.vicomtech.es/ingles/html/proyectos/index_proyecto82.html.

5. Conclusions and further work

We have presented a new method for moving people counting using a single camera. The algorithm is based on multiple line analysis. Our system is easily installable and doesn't need any special background as other systems need physical lines on the ground.

Our system works in complex situations, with groups of people crossing continuously going in and out at the same time. To measure the performance of the system under different circumstances three tests have been conducted. A correct people counting rate of 95% has been obtained.

In order to increase the precision of our system, we think that correcting the image distortion produced by the lens, could help us to decrease the number of false negatives near to the image edges. Also we want to make a more robust motion detection to reduce the same type of errors and to be able to work with hard

shadows, increasing the number of environments where our system could be installed.

Acknowledgments

This research is part of COPAF project in collaboration with Lurbe Grup S.A and Zeuxa Solutions S.L. The project is funded by the Basque INTEK programme.

References

- [1] L. Snidaro, L. Micheloni and C. Chiavedale, "Video Security for Ambient Intelligence", *IEEE Tr. on Systems, Man and Cybernetics*, Jan 2005, Vol. 35 N. 1, pp. 133-144.
- [2] P. Kilambi, E. Ribnick, A. Joshi, O. Masoud and N. Papanikolopoulos. "Estimating Pedestrian Counts in Groups", *Int. J. C. V. and Image Understanding*, 2007.
- [3] I. Haritaoglu, D. Harwood and L. S. David, "W4: Real-Time Surveillance of People and Their Activities", *IEEE Tr. on P. A. and Machine Intelligence*, Aug. 2000, pp. 809-830.
- [4] X. Zhang and G. Sexton, "A new method for pedestrian counting", *Fifth Int. Conf. on Image Processing and its Applications*, Edinburgh, UK, Jul 1995, pp. 208-212.
- [5] X. Liu, P.H. Tu, J. Rittscher, A. Perera and N. Krahnstoever, "Detecting and counting people in surveillance applications", *IEEE Conf. on Advanced Video and Signal Based Surveillance*, Italy, Sep. 2005, pp.306-311.
- [6] M. Han, W. Xu, H. Tao and Y. Gong, "An algorithm for multiple object trajectory tracking", *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Washington, USA, June 2004, pp. 864-871.
- [7] V. Rabaud and S. Belongie, "Counting Crowded Moving Objects", *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, New York, USA, June 2006, pp. 705-711.
- [8] G.J. Brostow and R. Cipolla, "Unsupervised Bayesian Detection of Independent Motion in Crowds", *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, New York, USA, June 2006, pp. 594-601.
- [9] A. Albiol, "Real-Time High Density People Counter Using Morphological Tools", *IEEE Int. Conf. on Pattern Recognition*, Washington, DC, Sept. 2000, Vol. 4, pp. 4652.
- [10] J. Bescós, J.M. Menéndez and N. García, "DCT based segmentation applied to a scalable zenithal people counter", *Int. Conf. on Image Processing*, Barcelona, Spain, Sept. 2003, pp. 1005-1008.