Second Asian Conference on Computer Vision (ACCV95), Singapore, 5-8 December, Vol. III, pp. 6-10 (invited)

IMAGE PROCESSING TECHNIQUES FOR CROWD DENSITY ESTIMATION USING A REFERENCE IMAGE

Jia Hong Yin, Sergio A. Velastin and Anthony C. Davies

Vision and Robotics Laboratory
Department of Electronic and Electrical Engineering
King's College London, Strand, London WC2R 2LS, UK
j.yin@kcl.ac.uk, s.velastin@kcl.ac.uk, tonydavies@kcl.ac.uk

Abstract

There are still significant problems in the planning, design and management of public facilities subject to dense pedestrian traffic. The automation of data collection and analysis of crowd behaviour is increasingly desirable in design of facilities and long-term site management using image processing techniques with existing closed-circuit television systems. We have investigated a number of techniques for crowd density estimation, movement estimation, incident detection and their relative merits using image processing. This paper presents techniques for background generation and calibration to enhance the previously-developed method of crowd density estimation using a reference image. An intensity region related to the average pixel intensity of each image in a sequence of crowd images is used to segment background pixels for generating a background image without pedestrians. The calibration approach, with which a previously-established relationship between image parameters and crowd density at one site can be used to estimate crowd density at various sites, involves calibration of the crowd image as opposed to calibration of the camera. Both techniques may be used in other surveillance systems such as vehicle monitoring.

1. Introduction

Crowd data such as density and flow are an important factor in the planning, design and management of public facilities subject to dense pedestrian traffic. It is well-known that conventional manual measurement techniques are not suitable for comprehensive data collection of patterns of site occupation and movement in such situations and automation of data collection to assist in understanding crowd behaviour is highly desirable [1]. Previous studies [2, 3] have shown the feasibility of crowd data estimation using image processing techniques with existing closed-circuit television systems. This may be applied to improving design of the built environment and increasing pedestrian safety.

Our initial procedure for crowd density estimation is to establish relationships between image parameters from various image processing techniques and actual crowd densities at an investigated site [4]. One of the techniques developed is to classify the pixels occupied by pedestrians in crowd images by comparison with a reference image with only background. Clearly, the reference image quality in this method influences the accuracy of measurement. The direct way to obtain a background image is to capture one from the investigated site when no pedestrians are present. However, in typical public areas, there are almost always some pedestrians present, and direct acquisition of an image with background only is often impracticable. Furthermore, inaccurate estimation may result from using the manually captured reference image because of effects of subsequent change in lighting level. Therefore, it is desirable to have an automatic method to generate reference images with only background for crowd density measurement at any time or at various investigated sites.

Another problem of these methods is that the establishment of a relationship between image parameters and number of people has to be repeated for each new site, which is time consuming and seriously restricts the adoption of automated methods. For general applicability of the developed methods, a faster method is needed which can transfer the previously-established image parameter to crowd density relationship in one investigated site to any other site.

In this paper, the technique of using a reference image to estimate crowd density is reviewed first. This is followed by introducing an automatic method to generate a background image from a sequence of crowd images. A calibration approach, by which the established relationship between image parameters and crowd density at one site can be used to estimate crowd density at other sites or for other camera positions, is presented with experimental results.

2. Density estimation using a reference image

The hypothesis of our method of density estimation using a reference image is that the area occupied by pedestrians in a crowd image contains information about crowd density in the scene. A fast method has been developed to estimate crowd density using a reference image with only background to classify pixels in a crowd image by subtraction as belonging to either pedestrians or background. Figure 2 shows a result image from Figure 1 which is a typical crowd image from one investigated site called “Site1”.

Fig. 1: Image from “Site1”

Fig. 2: Result image from Fig. 1

By comparing the number of pedestrian-classified pixels and number of people counted manually, a clear tendency is found out that the number of pixels increases approximately linearly with increase of number of pedestrians, as shown in Figure 3. A linear function relating number of pedestrian-classified pixels (N_{px}) to number of people (N_p) is obtained by using a least squares fit, as shown as follows:

$$N_p = a N_{px} + b \qquad (1)$$

where *a=0.00047* and *b=-2.81* for “Site1”. The standard deviation is 1.1 pedestrians. This linear function has been shown adequate for moderate crowd densities (e.g. up to 1.25 pedestrians/m^2 [5]).

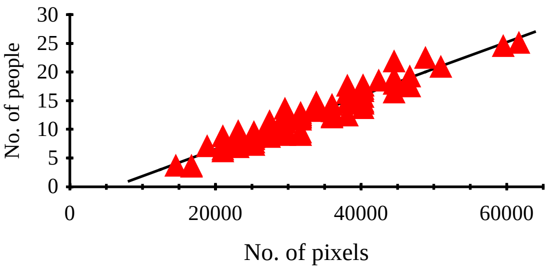

Fig. 3: No. of people vs. No. of classified pixels

3. Background generation

The technique using a reference image to classify pedestrian pixels has clearly shown its feasibility for crowd density measurement. However, the applicability and accuracy of this method obviously relies on the reference image. Apart from the inconvenience of capturing such a reference image without pedestrians in public facilities, accuracy deteriorates as a result of subsequent changes in lighting level and direction. Therefore, an automatic method to generate background images for this density estimation method is desirable.

Rodriguez [6] has obtained good results on image segmentation by successive background extraction, but he required local homogeneous background images, and he did not study the removal of objects from background. The contributions from Long [7] include *Smoothness Detector Method*, *Adaptive Smoothness Detector Method* and *Stable Fill Method* to detect background using a sequence of images with a moving object. The first produced good results in some views but limitations occur when the object was large and moved slowly. Many iterations to process the sequence are used in the latter two methods, which is time consuming and is not practical in surveillance systems. Our approach, to be described, aims to automatically generate background images from crowd video sequences, which are used as reference images for crowd density measurement.

3.1 Methodology

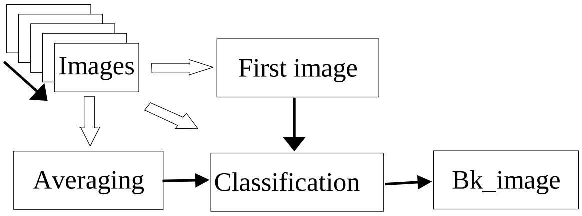

Fig. 4: Background generation procedure

The procedure for background generation is shown in Figure 4. A region with intensities between $\overline{A} - \Re$ and $\overline{A}$ is selected and used to classify pixels as background in each image of the video sequence. $\overline{A}$ denotes the average pixel intensity in each image, and $\Re$ is related to intensity variation of background pixels and is found by trial. Pixels in each image with intensities between $\overline{A} - \Re$ and $\overline{A}$ are classified as background pixels and extracted into a background frame store. This process is repeated for each image in the sequence, exploiting the fact that as people move in the sequence, more portions of the background are exposed so that for those parts where background pixels cannot be obtained from the first image they may be captured from the next. Any parts whose background pixels cannot be captured throughout the whole sequence are allocated an intensity corresponding to their value in the first crowd image. This assumes that these unclassified parts must belong to a part of the image which does not change throughout the whole sequence.

3.2 Analysis of results

Fig. 5: Generated background image

This method has been used to generate background images in our investigated sites for use as reference images to estimate crowd density in both data gathering and “incident” detection. One generated background image for “Site1” is shown in Figure 5. The results from this approach are adequate for use in our applications. However, results may be poor if the intensity of pedestrian clothing is similar to the background. Also for those parts of the image where pedestrians never move throughout the sequence, the background cannot be recovered.

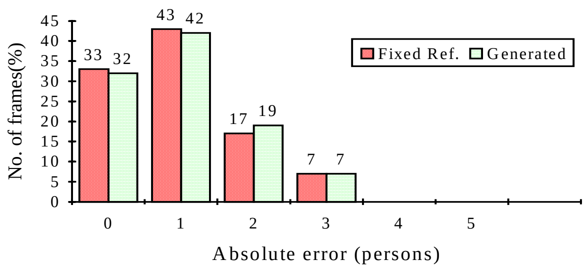

Fig. 6: Absolute error

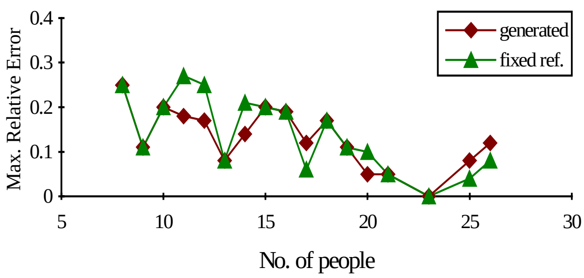

Fig. 7: Maximum relative error

The generated background image shown in Figure 5 has been used to estimate crowd density at “Site1”. Results using both a generated image and a reference image taken from the site have been compared with manual counts. The distribution of the *Absolute Error* (defined as the absolute difference between automatic counts and manual counts for each image) is shown in Figure 6. Figure 7 shows the *Maximum Relative Error* (defined as the ratio of maximum absolute error to the manual count). The results from using the generated reference image are very close to those using a reference image taken from the site.

4. A calibration approach for density measurement at various sites

Equation (1) can be used to measure crowd density but only for the site and the camera position at which the slope *a* and intercept *b* are obtained. Using this function to estimate crowd density at other sites (or other camera positions) is possible only if a relationship between the parameters from “Site1” and those applicable to each other site or camera position can be established. Some existing techniques in 3-D computer vision for camera calibration involve computing intrinsic and extrinsic camera parameters based on knowledge of 3-D coordinates, such as using a number of points [8, 9] or various targets [10], and their corresponding 2-D image coordinates. However, to avoid the complexity inherent in such methods, the calibration approach proposed here is intended to calibrate the crowd image as opposed to calibrating camera parameters. It exploits environmental constraints with which the relationship established at “Site1” can be used to estimate crowd density for other camera positions or for other sites.

4.1 Formulation

Using a conventional simple camera model and from the geometry of two similar triangles, the height h of an object in an image plane is approximately linearly proportional to its height H in 3-D world, and linearly proportional to its image position x as the object moves within the camera view [11]. If two objects of heights H_1 and H_2 placed at the same point are viewed using camera position one, and their corresponding heights in the image plane are h_1 and h_2, then $h_1 / h_2 = H_1 / H_2$. If the two objects are then viewed from a different camera position, called position two, then $h_1^* / h_2^* = H_1 / H_2$ where * denotes their heights in the second image plane. Therefore the heights in the two image planes are directly related as follows:

$$h_1 \cdot h_2^* = h_1^* \cdot h_2 \qquad (2)$$

Crowd density is measured in terms of “nominal” (average) pedestrians. For the purpose of calibration, a nominal pedestrian can be described by circumscribing rectangles, located near and away from the camera (Figure 8) both in the known image (“Site1”) and the image to be calibrated (“Site2”). The process can be based on images acquired using calibrating objects. When this is not possible, the nominal rectangles can be estimated manually.

Suppose that the rear rectangle at x_1 in the image plane of "Site1" has height h_1 and width w_1 and the front rectangle at x_2 has height h_2 and width w_2. The following equations can be derived from the linear relationship between image height *h* and its position *x*:

$$h = A_{Ih} x + B_{Ih} \tag{3}$$

where $A_{Ih} = \frac{h_2 - h_1}{x_2 - x_1}$ and $B_{Ih} = \frac{h_1 x_2 - h_2 x_1}{x_2 - x_1}$.

$$w = A_{Iw} x + B_{Iw} \tag{4}$$

where $A_{Iw} = \frac{w_2 - w_1}{x_2 - x_1}$ and $B_{Iw} = \frac{w_1 x_2 - w_2 x_1}{x_2 - x_1}$.

Let $s_I(x)$ indicate the area $h \cdot w$ of the rectangle in the image plane of "Site1", which can be described by equation (5):

$$s_I(x) = S_{I1} x^2 + S_{I2} x + S_{I3} \tag{5}$$

where $S_{I1} = A_{Ih} A_{Iw}$, $S_{I2} = A_{Ih} B_{Iw} + A_{Iw} B_{Ih}$ and $S_{I3} = B_{Ih} B_{Iw}$.

Similarly the area $s_{II}(x)$ of the rectangle in image plane of "Site2" is given by:

$$s_{II}(x) = S_{II1} x^2 + S_{II2} x + S_{II3} \tag{6}$$

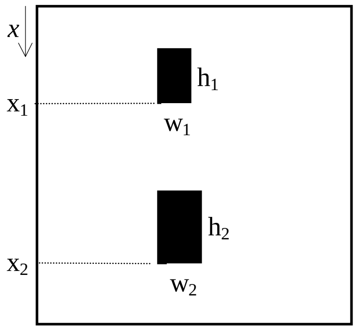

Fig. 8: Image for calibration

Based on the concept of nominal pedestrian size, if the number of pedestrian-classified pixels at *x* in a "Site2" image is $N_{II}(x)$, the number of corresponding pixels at *x*, $N_I(x)$, in a "Site1" image can be calculated by equation (7) in accordance with equation (2).

$$N_I(x) = \frac{s_I(x)}{s_{II}(x)} N_{II}(x) = \delta(x)\, N_{II}(x) \tag{7}$$

where $\delta(x) = \frac{S_{I1} x^2 + S_{I2} x + S_{I3}}{S_{II1} x^2 + S_{II2} x + S_{II3}}$.

The total number of pedestrian-classified pixels in an area of interest (AOI) in the "Site2" image can be equivalently converted to the total number of pixels in the "Site1" image with the following equation:

$$N_{px} = \underset{x \in AOI}{\quad} \delta(x) N_{II}(x) \tag{8}$$

which can be substituted into equation (1) to calculate the number of people in the area of interest in the "Site2" image.

The procedure of estimating crowd density using this calibration approach involves the following steps: $\delta(x)$ is first obtained by using the nominal rectangle located in "Site1" (the camera position where the function between number of pedestrian-classified pixels and number of people is known) and in "Site2" (where the measurement of crowd density is required). Then, the number of corresponding pedestrian-classified pixels is calculated using equation (8); finally this number is converted to number of people at the investigated site using equation (1).

4.2 Experimental results

Fig. 9: Image from "Site2"

Fig. 10: Image from "Site3"

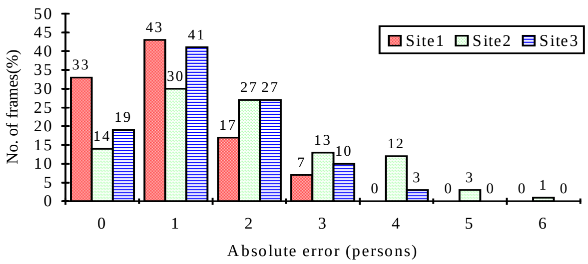

Fig. 11: Absolute error

This calibration approach has been used experimentally to estimate crowd density in two sites, called "Site2" and "Site3", together with equation (1) obtained from "Site1". $\delta(x)$ was obtained by measuring the area covering pedestrians in the images. Two background images were generated using the automatic method for the two sites. At "Site2", shown in Figure 9, large crowds of people had built up waiting for trains. Figure 10 shows a typical image from "Site3", a view in front of escalators. The results from the two sites have been analysed by comparing those obtained from "Site1" using a reference image taken directly in the site. Figure 11 and Figure 12 show the comparison of *Absolute Error* and *Maximum Relative Error* respectively. Results from "Site2" have errors larger than those for "Site3" due to over-crowding and poor lighting condition in the evening. The results from both sites are good enough for the purposes of automatic crowd monitoring and incident detection.

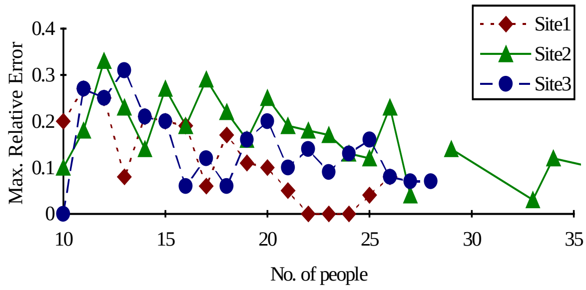

Fig. 12: Maximum relative error

5. Conclusions

This paper has presented two image processing techniques to enhance crowd density measurement using a reference image. A reference image with only background is used to classify image pixels as belonging to either pedestrians or background so that a functional relationship between number of pedestrian-classified pixels and number of people can be established for crowd density measurement. The automatic background generator proposed exploits an intensity region related to the average pixel intensity of each image to segment background pixels and puts background pixels from a sequence of crowd images together to obtain an artificial background image. The calibration approach, with which a previously-established function for one site can be used to estimate crowd density at various other sites, involves calibration of the image as opposed to calibration of the camera. The results from two investigated sites using the background generator and the calibration have been shown to be compatible with those from the site where the function between number of pedestrian-classified pixels and number of people was established with a reference image taken from site. The techniques increase the general applicability of the crowd-density estimation method developed by the authors. Both background generation and calibration may be used in other surveillance systems such as vehicle monitoring.

6. Acknowledgements

The authors are grateful to the UK EPSRC for its financial support (grants GR/H78511 and GR/83539), to Railtrack East Anglia (UK) for granting access to their sites, and to their colleagues X. Zhang, M. A. Vicencio-Silva, R. E. Allsop, L. Vaughan and A. Penn (University College London), for their collaboration.

7. References

[1] Davies A.C., Yin J.H. and Velastin S.A. (1995): "Crowd monitoring using image processing", IEE Electronics Communication Engineering Journal, Vol. 7, No. 1, February, pp 37-47.

[2] Velastin S.A., Yin J.H., Davies A.C., Vicencio-Silva M.A., Allsop R.E. and Penn A. (1994): "Automatic Measurement of Crowd Density and Motion Using Image Processing", IEE 7th International Conference on Road Traffic Monitoring and Control, 26-28 April 1994, London, UK, pp 127-132.

[3] Velastin S.A., Yin J.H., Davies A.C., Vicencio-Silva M.A., Allsop R.E. and Penn A. (1994): "Image Processing for On-line Analysis of Crowds in Public Areas", 7th IFAC/IFORS Symposium on Transportation Systems: Theory and Application of Advanced Technology. 24-26 August, Tianjin, China, pp 170-175.

[4] Yin J.H., Velastin S.A. and Davies A.C. (1994): "Measurement of Crowd Density Using Image Processing", EUSIPCO'94. September, Edinburgh, UK, pp 1397-1400.

[5] Polus A., Schofer J. L. and Ushpiz A. (1983): "Pedestrian Flow and Level of Service", Journal of Transportation Engineering, Vol. 109, No. 1, January, pp 46-56.

[6] Rodriguez A. A. and Mitchell O. R. (1991): "Image Segmentation by Successive Background Extraction", Pattern Recognition, Vol. 24, No. 5, pp 409-420.

[7] Long W. and Yang Y. (1990): "Stationary Background Generation: An Alternative to the Difference of Two Images", Pattern Recognition, Vol. 23, No. 12, pp 1351-1359.

[8] Tsai R.Y. (1987): "A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, August, pp 323-344.

[9] Wei G.Q., Ma S.D. (1994): "Implicit and Explicit Camera Calibration: Theory and Experiments", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 5, May, pp 469-480.

[10] Wang L.L. and Tsai W.H.(1991): "Camera Calibration by Vanishing Line for 3-D Computer Vision", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 4, April, pp 370-6.

[11] Yin J.H., Velastin S.A. and Davies A.C. (1995): "A Calibration Approach for Estimation of Crowd Density Using Image Processing", accepted for 2nd International Workshop on Image and Signal Processing: Theory, Methodology, Systems and Applications, 8-10 November, Budapest, Hungary.