AUTOMATED MEASUREMENT OF CROWD DENSITY AND MOTION USING IMAGE PROCESSING

S.A. Velastin †, J.H. Yin †, A.C. Davies †, M.A. Vicencio-Silva ‡, R.E. Allsop ‡, A. Penn ‡

† King's College London, UK. ‡ University College London, UK

INTRODUCTION

The characteristics of pedestrian flow in built environments have been investigated for some years. The modelling of pedestrian movement patterns and their relationships with space configuration has been shown possible when based on manual observations of individual pedestrians, as reported by Hillier et al (1). However, because human observers have limited attention span and reliability, such data collection methods have proved to be difficult to apply to crowded conditions. Thus, significant problems still exist for those involved in designing, managing and policing urban areas subject to major crowd movements. This is a daily problem for certain classes of building such as station concourses, airports, stadia, etc. where manual observation is limited to qualitative situation assessment. Therefore, improved ways of automated data collection and processing are needed, using existing CCTV systems, leading to improved designs, management procedures, public safety and lower operating costs.

REVIEW OF PREVIOUS WORK

Manual Measurement Methods

A wide variety of factors are involved in determining pedestrian behaviour, including age, gender, physical fitness, social relationship to neighbouring pedestrians, purpose of journey, area topography, etc. There is a number of conventional techniques used at present. Direct in-site measurement by human observers is limited in practice to low densities or qualitative assessment. In controlled experiments, e.g. Hankin and Wright (2), a known number of subjects are asked to walk "naturally" in a given area, but the method is limited by the uncertainty of being able to reproduce natural behaviour and by the associated logistics problems. Time-lapse photography is extensively used, e.g. Navin and Wheeler (3), Fruin (4), combined with statistical models to derive pedestrian speeds and densities. Finally, off-line manual analysis of video recordings is also popular, e.g. Polus et al (5), Tanaboriboon et al (6), Tanaboriboon and Guyano (7). Typical conclusions from this type of work include:

- Males tend to walk faster than females.
- Speed is inversely related to density.
- Flow rates (volume) and velocities decrease with an increase in opposing pedestrian traffic.
- Speeds may vary considerably depending on the nature of the trip, time of day and weather.
- Pedestrians attempt to maintain "buffer zones" to prevent collisions with other pedestrians or obstacles. The size of these buffer zones normally increases with speed (possibly in a non-linear manner).
- Pedestrian behaviour in various countries may differ according to body structure and cultural conventions. Hence, the criteria for service standards in Europe may not be appropriate for the Far East and vice versa.

Some of these results have assisted in the design of new pedestrian facilities, such as subways, airport terminals, and underground stations, e.g. see Hoel (8), Davis and Braaksma (9), playing a significant role especially where space availability is limited and land expensive. However, the use of manual techniques is a major limiting factor for extending this kind of work to crowded conditions, long-term studies or for situations where on-line surveillance is required. Although human labour is accurate for short time intervals and difficult to replace for qualitative situation assessment, it is costly, slow and deteriorates when large amounts of data need to be analysed. Nevertheless, it is important to incorporate some of the above findings in any automatic system, as part of its heuristics and *a priori* knowledge.

Automatic Pedestrian Detection

Computer vision is a well-established field that has produced important advances in methods and systems. Hardware cost reductions in combination with a large installed base of CCTV monitoring systems in public areas has led to a growing interest on the application of image processing techniques for the detection and analysis of pedestrian traffic. For example, Ishii et al (10) devised a technique to measure bi-directional pedestrian flow from plan view images. Khan and Ince (11) have shown how crowd densities can be estimated from digitised aerial photographs by measuring the number of edge pixels in such images. Plan views are difficult to obtain in built urban environments (especially if existing CCTV systems are used), but this paper shows that density information can be extracted from edge features even for these cases. Rourke and Bell (12) have proposed a technique for estimating occupancies based on inter-frame image differences showing real-time potential with an implementation on a transputer network. It is not clear if the approach has been tested for crowded conditions. Finally, Bartolini et

Road Traffic Monitoring and Control, 26-28 April 1994, Conference Publication No. 391, © IEE, 1994

al (13) report a method to measure the number of people getting in an out of a bus, but the technique is limited to bi-directional flow of a reduced number of pedestrians. Therefore, most reported methods impose assumptions (such as bi-directional flow, plan views or low density levels) which significantly limit their applicability to crowded conditions in semi-confined areas using existing camera positions. A number of researchers have used approaches that aim at identifying and tracking individual pedestrians, from which crowd density and motion could be derived. These methods tend to have poor performance in the presence of occlusion. This paper shows that density and motion can be measured without explicit knowledge of individual shape, size and velocity.

SYSTEM OVERVIEW

The development environment consists of equipment for on-site recording and a general-purpose image processor hosted by a Sun workstation for real-time implementation and algorithmic development. Video recordings were made at peak times in a busy commuter railway station in London (UK), using two camera positions typical of conventional surveillance CCTV equipment. The video images were digitised (512x512 pixels, 256 grey levels), processed and stored on disk for further manual analysis (for comparisons and calibration). Fig. 1 shows a typical digitised image with about 17 people within an "area of interest" (AOI), shown by a white rectangle.

Fig. 1: Digitised (original) image

CROWD DENSITY

Two pixel-based methods have been devised to estimate crowd density, as described below.

Thinned Edges

A computationally efficient three-pixel-neighbourhood edge detector is applied to a single original image, based on the findings reported in (11), followed by a horizontal thinning operator to reduce the effects of different shapes, sizes, texture and intensity in pedestrians. A typical result is shown in Fig. 2. Pedestrian density has been found to be correlated to the number of thinned edge pixels in the processed images. The use of other features, such as gradient magnitude, has shown no significant improvements over simple pixel counts.

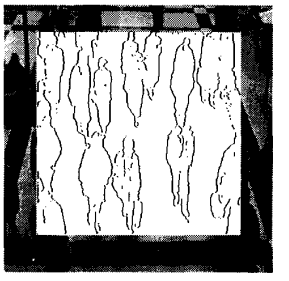

Fig. 2: Thinned Edges

To characterise this method, long image sequences have been digitised ensuring that no two consecutive images contain the same pedestrians, by extending the interframe interval to typically 10 seconds. Fig. 3 shows the relationship between the number of thinned edge pixels and manually counted pedestrians for a sequence of approximately 100 images. The difference between the measured data and a straight line approximation obtained by a least squares fit has a standard deviation of 1.69 pedestrians. This is equivalent to a 2σ relative error of 23% for 15 pedestrians, the level at which crowding starts for the observed area, according to the classification proposed in (5). A logarithmic least squares fit gives similar results.

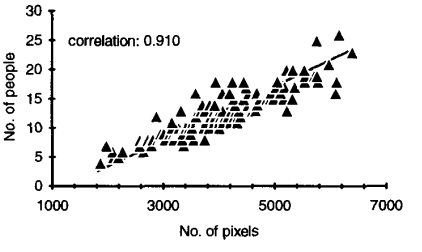

Fig. 3: No. people vs. No. pixels (thinned edges)

Background Removal

This method is based on subtraction from a background-only image (Fig. 4) which, ideally, identifies the pixels occupied by pedestrians. A three-pixel-neighbourhood is used to reduce noise. Fig. 5 shows a typical image after background removal.

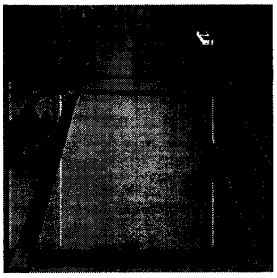

Fig. 4: Background image

Fig. 5: Background removal

Fig. 6 shows the relationship between the number of pixels after background removal and manually counted pedestrians. The difference between the measured data and the best fit has a standard deviation of 1.12 pedestrians, equivalent to a 2σ relative error of 15% for 15 pedestrians. A logarithmic least squares fit gives similar results.

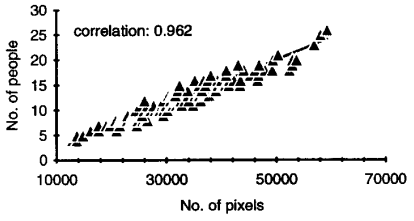

Fig. 6: No. people vs. No. pixels (background removal)

Measurement errors (a combination of quantisation errors, sensor noise, illumination changes, occlusion, etc.) in these methods have been observed to be practically zero-mean. Thus, they can be reduced by averaging results over a number of images. This is feasible since these methods can operate at near real-time video rates on most general-purpose image processors. Fig. 7 shows a typical result for background removal after averaging over ten images. In this case, the standard deviation has been reduced to 0.46 pedestrians, a 2σ relative error of 6% for 15 pedestrians. This level of accuracy is accepted as satisfactory for monitoring purposes and for some data gathering applications.

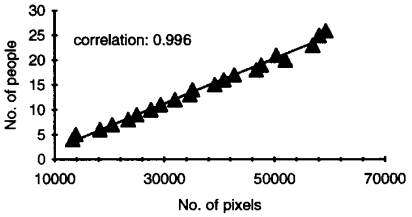

Fig. 7: Effect of averaging (background removal)

Kalman Filtering

The authors have proposed the use of a Kalman filter, Brown and Hwang (14), to integrate the two independent measuring techniques presented above. The approach can exploit the high data rates of video images compared with the rates required for data gathering and monitoring. The use of a simple dynamic model that assumes no change in pedestrian density from one image to the next, other than zero-mean modelling noise to account for real changes in density, results in a mean relative error of less than 8% when operating at an interframe interval of 10 seconds. For further details see Velastin et al (15).

Geometric Correction

In the methods described above the relationship between number of pixels, or any other similar image feature, and the number of pedestrians depends on camera position and hence can be obtained through an initial calibration procedure. This is not a significant problem as cameras are rarely moved once installed. More importantly, because a global measure is used, the methods implicitly assume a relatively homogenous pedestrian distribution. Perspective distortions can introduce errors if, for example, flow is concentrated on a particular part of the image. To reduce these effects, the images can be geometrically corrected such that all pedestrians have similar sizes, as shown in Fig. 8.

Fig. 8: Geometrically corrected image

A system developed in King's College London by Papadopoulos and Clarkson (16, 17) has been used. Accuracy is similar to that obtained without correction, as the process adds no further information.

HEAD MOVEMENTS

A technique for incident detection used by many on-line crowd monitoring operators consists in checking for the presence of vertical head or body oscillations. This is particular useful when there is a large number of TV displays, for large levels of crowding or low height cameras (e.g. underground stations). The absence of such oscillations in conditions where continuous flow is expected is an indication of a stationary crowd and hence of a potentionally dangerous situation. In collaboration with the authors, Hentschel (18) investigated frequency-domain techniques to identify these vertical oscillations for discriminating between stationary and non-stationary flow, proposing the Linear Area Transform (LAT) algorithm. In a running image sequence I_0, I_1, .. I_m, corresponding pixels are subtracted in consecutive images to obtain an *interframe* sequence (of moving pixels) D_0, .. D_{m-1}. A single value is computed for each D_k in the sequence, as follows

$$g_k = \sum_{i=0}^{N-1} i \sum_{j=0}^{N-1} \mathbf{D}_{i,j} \quad , \quad k = 0,1,\ldots,m-1 \qquad (1)$$

where NxN is the image size, i is row number and j is column number. In other words, the total amount of motion in each image, weighted by vertical position, is accumulated in g_k. The Discrete Fourier Transform of $g_k - g_{k-1}$ for the sequence is then calculated. The frequencies of head oscillations correspond to peaks in the resulting frequency spectrum. Fig. 9 shows a typical result for a sequence with 32 images.

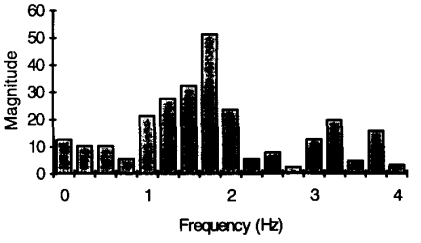

Fig. 9: LAT Frequency spectrum

The peak at 1.75 Hz matches manual measurements. The LAT compares favourably with conventional methods such as the Cosine Area Transform.

MOTION ESTIMATION

In semi-confined spaces individuals are free to move in various directions. Moreover, at any one instant, different parts of a single individual (e.g. head, limbs) move in different ways. However, crowd analysis is usually more concerned with group behaviour such as preferential motion direction and magnitude. For instance, a typical useful measure is the distribution of the proportion of people moving in a discrete set of preferential directions (e.g. a "wind rose"). The approach taken by the authors is to measure motion features at pixel or pixel-neighbourhood level which are then aggregated to obtain motion properties for larger regions in an image. The aggregated results can then be used to establish overall preferential crowd velocities (direction and magnitude).

Conventional Optical Flow Computation

A well-known method to measure motion is the optical flow computation technique proposed by Horn and Schunck (19). The authors have found that improved results and reduced computation cost are obtained if both images are first subtracted from a background image to preselect pixels of interest. Small disjoint neighbourhoods are used to average the vector field. Fig. 10 shows a typical result when processing an image containing a standing queue (right hand side) and moving pedestrians (left hand side). The average optical flow vectors have been superimposed on the original image (the small white square indicates the origin of each vector).

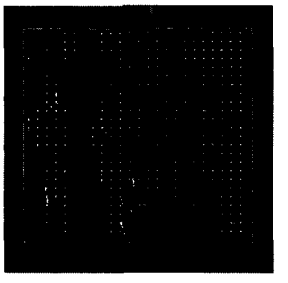

Fig. 10: Optical flow (19)

This simple approach is reasonably successful at computing directions but, as has been reported elsewhere, magnitudes are not reliable when they exceed two or more pixels/frame. This is a serious limitation for real-time implementation (as it imposes the need for high frame rates) and therefore the use of this method is limited to discrimination between stationary and non-stationary crowds.

Motion by Frame-to-Frame Intensity Correlation

To overcome the problems associated with the conventional optical flow calculation algorithm, the authors devised a different technique based on grey-level correlation of small neighbourhoods in two consecutive images, herein called F and S. To reduce the amount of data to process, and hence the computation time, the first image (F) is segmented (i.e. *crowd* pixels are extracted) by one of three methods: background removal, thinned edges detection or interframe pixel motion (e.g. pixels in the difference frame $|F-S|$ which are above a threshold). Then for each segmented pixel in F a small pxp neighbourhood is defined. The hypothesis is that this neighbourhood represents a small pedestrian region that does not change significantly in shape, size or intensity from one image to the next. The displacement of this area is found by an exhaustive search amongst all possible pxp regions in a larger sxs neighbourhood in the second image (S), centred at the same pixel position. The search computes a matching function, defined as follows:

$$m(is,js) = \sum_{ip=-\frac{p}{2}}^{\frac{p}{2}} \sum_{jp=-\frac{p}{2}}^{\frac{p}{2}} \left| F_{ic+ip,jc+jp} - S_{ic+ip+is,jc+jp+js} \right| \qquad (2)$$

$$is, js = -\frac{s}{2} \ldots \frac{s}{2}$$

where (ic, jc) are the coordinates of the segmented pixel and (is, js) are the centre coordinates of the "candidate" matching pxp area within the sxs region in the second image. The displacement vector is calculated from,

$$m(is', js') = \min(m(is, js)) \qquad (3)$$
$$\Delta i = is' - ic, \Delta j = js' - jc$$

The result is an image velocity vector field defined for each segmented pixel. Figs. 11, 12 and 13 show typical results for segmentation based on background removal, frame difference and thinned edges respectively.

Fig. 11: Background removal preselection (p=5, s=11)

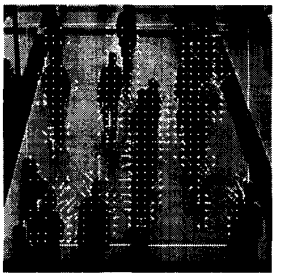

Fig. 12: Interframe difference preselection (p=5, s=11)

Fig. 13: Thinned Edges preselection (p=5, s=11)

To aid visualisation, these vectors have been averaged in disjoint neighbourhoods and superimposed on the first image. Magnitudes are more reliable than those obtained with the Horn algorithm, even for displacements of over 10 pixels.

It is difficult to assess the accuracy of these algorithms compared with manual methods and to visualise motion tendencies. Thus the data is aggregated on polar plots where the direction angles are divided into a discrete range (e.g. $\Delta\theta = 1°$). All the velocity vectors within a given range are added and the results presented in a polar r-θ plot. Figs. 14, 15 and 16 show the results obtained for the images in Figs. 11, 12 and 13 respectively. The main "south east" tendency can now be clearly seen. The results obtained with background removal are similar to those with frame difference. Segmentation based on thinned edges is poor in comparison, showing a larger degree of dispersion.

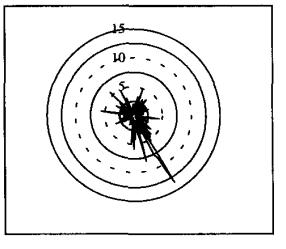

Fig. 14: Motion r-θ plot (background removal)

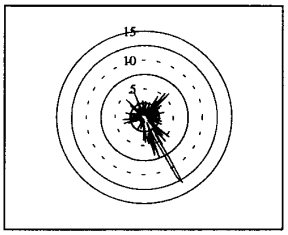

Fig. 15: Motion r-θ plot (frame difference)

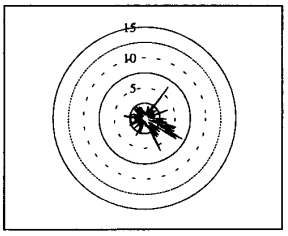

Fig. 16: Motion r-θ plot (thinned edges)

CONCLUSIONS

Some image processing techniques applicable to the measurement of density and motion in crowded scenes have been presented. These give results comparable to human observers with potential for real-time implementation. A useful characteristic of these techniques is their independence from the apparent shape of individual pedestrians. Detailed studies of site usage and real-time monitoring then become feasible.

ACKNOWLEDGEMENTS

The work reported here was supported by UK SERC grants GR/H78511 and GR/83539. The authors are grateful to Network SouthEast and London Underground for providing video tapes and access to their sites and to Mr. Xu Zhang (Centre for Transport Studies, University College London) for his suggestions on motion detection.

REFERENCES

1. Hillier B., Penn A., Hanson J., Grakewski T., and Xu J., 1993, Environment and Planning, 20, 29-66
2. Hankin B.D. and Wright R.A., 1958, Operational Res. Quarterly, 9, 81-88
3. Navin F.P.D. and Wheeler R.J., 1969, Traffic Eng., June, 30-36
4. Fruin J.J., 1971, Highway Res. Record 355, 1-14
5. Polus A., Schofer J.L. and Ushpiz A., 1983, J. Transportation Eng., 109, 46-56.
6. Tanaboriboon Y., Hwa S.S. and Chor C.H., 1986, J. Transportation Eng., 112, 229-235
7. Tanaboriboon Y. and Guyano J.A., 1989, J. Inst. Transportation Eng., 59, 39-41
8. Hoel L.A., 1968, Traffic Eng., Jan., 10-13
9. Davis D.G. and Braaksma J.P., 1988, Transportation Research Part A, 22A, 375-388
10. Ishii H., Ono T., Takusagawa J. and Muroi N., 1987, J. Illuminating Eng. Inst. Japan, 71, 626-631
11. Khan M.A. and Ince F., 1989, Arabian J. of Science and Eng., 14, 541-549
12. Rourke A. and Bell M.G.H., 1992, "Video Image Processing Techniques and their Application to Pedestrian Data-collection", Research Report No. 83, Transport Operations Research Group, University of Newcastle upon Tyne, UK
13. Bartolini F., Capellini V., and Mecocci A., 1994, Image and Vision Computing, 12, 36-41
14. Brown R.G. and Hwang P.Y.C., 1992, "Introduction to Random Signal Analysis and Kalman Filtering", Wiley, 2nd ed.
15. Velastin S.A., J.H. Yin, A.C. Davies, M.A. Vicencio-Silva, R.E. Allsop R.E., and A. Penn, 1993, "Analysis of Crowd Movement and Densities in Built-up Environments using Image Processing", IEE Coll. on Image processing for Transport Applications, 9 Dec., London, UK. Digest No. 1993/236, 8/1-8/6
16. Papadopoulos C.A., and Clarkson T.G., 1992, Electronic Letters, 28, 2314-2315
17. Papadopoulos C.A., and Clarkson T.G., 1991, "Parallel Processing of digital images using a modular architecture", IEE 6th Int. Conf. on DSP in Communications, Sept., 99-104
18. Hentschel T., 1993, "Image Processing Techniques for the Estimation of Features of Crowd Behaviour in Urban Environments", MSc. Dissertation, King's College London, UK
19. Horn B.K.P. and Schunck B.G., 1981, Artificial Intelligence, 17, 185-203