ESTIMATION OF CROWD DENSITY USING IMAGE PROCESSING

A. N. Marana[1], S. A. Velastin[2], L. F. Costa[3] and R. A. Lotufo[4]

Abstract

Human beings perceive images through their properties, like colour, shape, size, and texture. Texture is a fertile source of information about the physical environment. Images of low density crowds tend to present coarse textures, while images of dense crowds tend to present fine textures. This paper describes a new technique for automatic estimation of crowd density, which is a part of the problem of automatic crowd monitoring, using texture information based on grey-level transition probabilities on digitised images. Crowd density feature vectors are extracted from such images and used by a self organising neural network which is responsible for the crowd density estimation. Results obtained respectively to the estimation of the number of people in a specific area of Liverpool Street Railway Station in London (UK) are presented.

1 Introduction

The management and control of crowds is a crucial problem for human life and safety [1, 2, 3]. Two important aspects of the problem of correct management and control of crowds are the design of the environments in which crowds are expected to arise and the real-time monitoring of crowds within existing, typically urban, structures.

The development of models of crowd behaviour can help architects and town-planners to design safer buildings. Sime [4] reviews crowd psychology in terms of its relationship to engineering and crowd safety and stresses the need to validate computer simulations of crowd movement and escape behaviour against psychological as well as engineering criteria.

For the problem of real-time crowd monitoring there is an established practice of using extensive closed-circuit television systems. However, the large number of video cameras often used by such systems requires a huge recording storage capacity and a number of people to observe the television monitors. For reasons of cost-effectiveness, only parts of the recordings are usually stored and most of times only one human

[1]DEMAC,IGCE, UNESP, Rio Claro, SP, Brazil.
[2]EEE, King's College London, London, UK.
[3]IFSC, USP, São Carlos, SP, Brazil.
[4]DCA, FEEC, UNICAMP, Campinas, SP, Brazil.

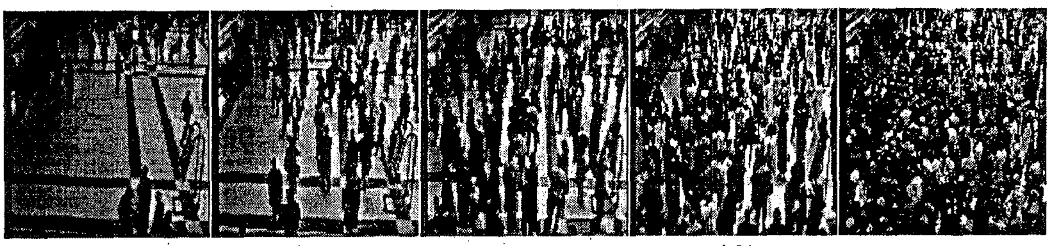

Figure 1: Images from Liverpool Street Railway Station, London, UK. (a) Very low density; (b) Low density; (c) Moderate density; (d) High density; (e) Very high density.

observer is responsible for the monitoring of many different areas through a two-dimensional array of monitors. As routine monitoring is tedious, the observers are likely to lose concentration. The advantages and necessity of automatic surveillance for routine crowd monitoring are, therefore, clear.

Automatic estimation of crowd density is a part of the problem of automatic monitoring of crowds. This paper describes a novel technique to estimate crowd density based on texture information of digitised images of the area under monitoring.

Figure 1 presents images of crowds ranging from very low to very high densities. Images of high density crowded areas are often made up of fine patterns, while images of low density crowded areas are mostly made up of coarse patterns (especially when the backgrounds are characterised by low frequencies). The technique described in this paper uses this fact to estimate crowd densities by using the grey level dependence matrix (GLDM) method [5] to carry out texture analysis and a neural network, implemented according to the Kohonen's self organising map (SOM) model, for the task of crowd density estimation.

Sections 2 and 3 review the GLDM method for texture analysis and the Kohonen's neural network SOM model, respectively. Section 4 describes the new technique for automatic crowd density estimation. Section 5 presents results obtained when this technique was applied to estimate crowd densities in real images from Liverpool Street Railway Station in London, UK. Conclusions are presented in section 6.

2 Grey Level Dependence Matrix

The grey level dependence matrix method proposed by Haralick [5] is based on the estimation of second-order joint conditional probability density functions, $f(i,j|d,\theta)$. Each $f(i,j|d,\theta)$ is the probability of the pair of grey levels (i,j) occurring in a pair of pixels of the image given that these pixels are separated by a distance d along the direction θ. The estimated values form a two-dimensional histogram which can be written in matrix form, the so-called grey level dependence matrix (GLDM).

Because GLDMs must be estimated for each pair of parameters (d, θ), it is usually computationally necessary to restrict these parameters to a limited number of values. Normally, only four angles (0^o, 45^o, 90^o and 135^o) and only one distance (one pixel) are used.

For a given pair of parameters (d, θ), the histogram obtained for fine texture tends to be more uniformly dispersed than the histogram for coarse texture. Texture coarseness can be measured in terms of relative spread of histogram occupancy cells about the main diagonal of the histogram. In this work, only four of the spread indicators proposed by Haralick for texture measurements have been used:

Contrast:

$$S_c(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-j)^2 f(i,j|d,\theta)$$

Homogeneity:

$$S_h(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\frac{f(i,j|d,\theta)}{1+(i-j)^2}$$

Energy:

$$S_g(d,\theta) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} f(i,j|d,\theta)^2$$

Entropy:

$$S_p(d,\theta) = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} f(i,j|d,\theta) log f(i,j|d,\theta)$$

where L is the number of grey levels of the image.

3 Self Organising Map Neural Network

Self organising maps (SOM) are neural network models proposed by Kohonen [6] to implement non-linear projections from high-dimensional spaces X onto low-dimensional maps M. The two-dimensional mapping (frequently adopted in Kohonen's model) is a function $f : X \rightarrow M$, where $X \subset R^n$ and $M \subset R^2$, which assigns to each element $x \in X$ a pair $(i,j) \in M$. The elements m_{ij} of the map M, as well as the input data, are n-dimensional vectors which hold the values of the synaptic strengths of the neural network.

For a given input stimulus $x \in X$, the winner node $(i,j) \in M$ is determined by the following conditions: $\|m_{ij} - x\| = min\|m_{kl} - x\|$, for $0 \leq i,j,k,l \leq D$ (D is the size of M) and $(k,l) \in M$.

In the Kohonen's SOM neural network model, the neurons in the output layer are connected to every cell in the input space and have lateral interactions, which are defined by a function $h_{ij}(t)$ of the distance between the winner node (i,j) and any other node (k,l) in the winner node's neighbourhood. Usually, a Gaussian function is used to determine the spatial influence amongst the nodes around the winner node.

The learning is unsupervised and carried out in the training stage according to the following incremental adjustments in the synaptic weights: $m_{kl}(t+1) = m_{kl}(t) + h_{ij}(t)[x - m_{kl}(t)]$, where $h_{ij}(t) = \alpha(t)exp\left[-\frac{\|r_{kl}-r_{ij}\|^2}{\delta(t)^2}\right]$, (i,j) identifies the winner node for the stimulus x, $\|r_{kl} - r_{ij}\|$ is the spatial distance between the winner node and the node (k,l) belonging to the winner node's neighbourhood N_{ij}, and $\delta(t)$ is the Gaussian's variance which defines the interaction level around the winner node. The function $N_{ij}(t)$, the learning rate $\alpha(t)$ and the Gaussian's variance $\delta(t)$ are adjusted during the training stage as a function of time.

After the training stage the network nodes are labelled by presenting a number of input vectors with known classification and assigning the nodes to different classes by majority voting, according to the frequency with which such nodes are closest to the calibration vector of a particular class.

4 Technique for Crowd Density Estimation Based on Texture Analysis

Figure 2 shows how images have been used in the proposed technique to estimate the number of people in areas under surveillance. The estimation of crowd densities is based on texture measures of the images and given in terms of ranges such as very low, low, moderate, high and very high densities. The number of people for each range and the number of ranges itself depend on the specific application and particular characteristics of the area being monitored.

The first step of the technique consists in to calculate measures of contrast, homogeneity, energy and entropy for four grey level dependence matrices, which are obtained by applying the GLDM method on the input image with $(d,\theta) = (1,0^o)$, $(1,45^o)$, $(1,90^o)$ and $(1,135^o)$. Next, the 16 measurements obtained in the first stage are put into a vector, which is used as the crowd density feature vector of the input image. Finally, this feature vector is used by a SOM neural network to estimate the range of crowd density. The range of density is estimated as one of the five possibilities.

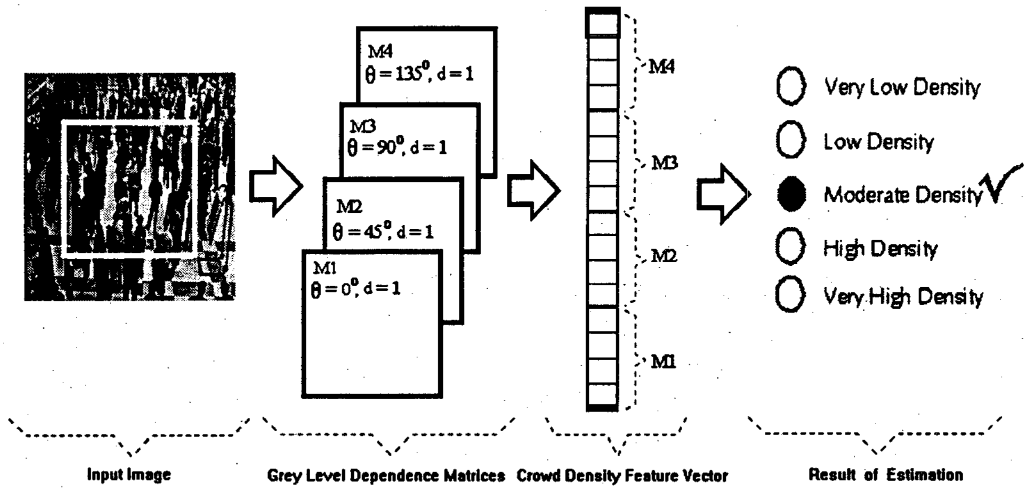

Figure 2: Crowd density feature vectors are extracted from the input image trough the texture analysis GLDM method.

Group	Train Set	Test Set	Label	Average	Standard Deviation	Correct Classification
Very Low	34	34	0	0.05	1.37	94.12%
Low	39	39	1	1.20	4.04	53.85%
Moderate	27	26	2	2.00	2.00	84.62%
High	19	18	3	3.05	0.97	94.44%
Very High	32	32	4	3.93	1.36	93.75%
Total	151	149	-	-	-	81.88%

Table 1: Results of automatic estimation of crowd density in an area of Liverpool Street Railway Station, London, UK, using image processing.

As the process of estimation of crowd density is supervised, it requires a previous stage where the neural network is trained. Therefore, it is necessary to collect images from the area of monitoring to be used as the training set of images for the neural network. Such a set must included enough number of samples for each one of the pre-defined ranges of crowd densities in order to be statistically significant. This estage requires a manual estimation of crowd density of the images of the training set.

5 Results

Table 1 shows results obtained when 149 images captured from an area of Liverpool Street Railway Station, London, Uk, were used to assess the accuracy of the new method. A set of 151 images was used as the training set for the neural network, summing up 300 images.

As well as the training set, crowd densities of images of the test set were manually estimated in advance, in order to establish a comparison standard. Based on the manual estimation, the test images were separated in groups of very low density (0-15 people), low density (16-30 people), moderate density (31-45 people), high density (46-60 people) and very high density (more than 60 people). The numbers of images in each train and test group are also presented in Table 1 (second and third columns, respectively).

Integer numbers from 0 to 4 were used to label the groups from very low to very high densities. These number are the values expected to be found for each group during the process of crowd density estimation (fourth column of Table 1). The means of estimation obtained for every groups, as well as its variances are presented in the fifth and sixth columns of Table 1.

The rightmost column present the rates of correctness for each group. It shows, for example, that the best result was obtained for the group of high crowd density images (94.44%), and the worst (53.85%), for the group of low crowd density images. The general rate of correct estimation reached 81.88%, that is, 122 images from a test set of 149 images were correctly estimated.

The regular result obtained for the group of low crowd density images can be better understood when it is observed that all the incorrect classification during this test were assigned to neighbours of the correct range of density, allowing the method to obtain mean estimations for each group very near to the expected values (the label numbers), with low variances.

6 Summary and Conclusion

This paper presents a new technique for the problem of automatic estimation of crowd density, a very important aspect of the problem of automatic crowd monitoring. This technique extracts crowd density features from digitised images of the area being monitored by using probabilities of grey-level transitions of such images. The crowd density features are used by a neural network to classify the crowd images according to five density classes.

Results obtained for images from an area of Liverpool Street Railway Station were used to assess the new method show that the proposed technique is really promising. Crowd densities of 122 images from a test set of 149 images were correctly estimated, that is, 81.88% of correctness. In addition, the groups of very low, high and very high crowd densities reached around 94% of correct estimation, a very good rate.

As all incorrectly classified images on this test were assigned to neighbours of the correct range of density, the committed error can be acceptable in some practical circumstances.

A previous technique proposed by Davies [11] for the same problem also allowed good performance. However, it can not be applied to images where there are superimposed people, which invalidates the linear relationship found to exist between the number of edge and foreground picture elements and the number of people in the image. For images with low crowd density, such technique can provide the exact number of people.

In contrast with the Davies' technique, the method presented in this paper is able to deal with high crowd density images, but its output is given in terms of ranges of crowd densities.

7 Acknowledgements

The authors are grateful to Railtrack PLC (London-UK), for granting access to their sites, and to Maria Alicia Vicencio-Silva, who first suggested the use of texture features to measure crowd densities. Aparecido Nilceu Marana is also grateful to CNPq (Proc.200823/95-7). Luciano da Fontoura Costa thanks FAPESP (Procs.94/3536-6 and 94/4691-5) and CNPq (Proc.301422/92-13) for financial help.

References

[1] Berlonghi, A. E. (1995) "Understanding and planning for different spectator crowds" *Safety Science*, Vol. 18, No. 4, 239-247.

[2] Dickie, J. F. (1995) "Major crowd catastrophes", *Safety Science*, Vol. 18, No. 4, pp. 309-320.

[3] Nicholson, C. E. (1995) "The investigation of the Hillsborough disaster by the Health and Safety Executive", *Safety Science*, Vol. 18, No. 4, 249-259.

[4] Sime, J. D. (1995) "Crowd psychology and engineering", *Safety Science*, Vol. 21, No. 1, 1-14.

[5] Haralick, R. M. (1979) "Statistical and Structural Approaches to Texture", *Proceedings of the IEEE*, Vol. 67, No. 5, pp. 786-804.

[6] Kohonen, T. (1990) "The Self-Organizing Map", *Proceedings of IEEE*, vol. 78, No. 9, pp. 1464-1480.

[7] Laws, K. I. (1980) "Textured Image Segmentation", University of Southern California, Image Processing Institute, USCIPI Report 940, January.

[8] Li, S. Z. (1995) "Markov Random Field Modelling in Computer Vision", *Springer-Verlag.*

[9] Raghu, P. P., Poongodi, R. and Yegnanarayana, B. (1995) "A Combined Neural Network Approach for Texture Classification", *Neural Networks*, Vol. 8, No. 6, pp. 975-987.

[10] Van Gool, L. (1985) "Texture Analysis Anno 1983", *Computer Vision, Graphics, and Image Processing*, Vol. 29, pp. 336-357.

[11] Davies, A. C., Yin, J. H. and Velastin, S. A. (1995) "Crowd monitoring using image processing", *Electronics and Communications Engineering Journal*, February, pp. 37-47.