"Evaluation of motion based algorithms for automated crowd management", Workshop on Performance Characterisation and Benchmarking of Vision Systems, International Conference on Vision Systems, ICVS, January 1999, Spain, pp. 80-96

Evaluation of motion-based algorithms for automated crowd management

B. A. Boghossian* and S. A. Velastin*

1 Introduction

Public safety and crowd management, are the significant reasons for covering more and more public areas with Closed Circuit Television (CCTV) cameras. These motives are encouraging governments and companies to invest large amounts of funds in CCTV equipment, their installation and their maintenance, overlooking the fact that the real value of CCTV systems lies in the experience and performance of the guards operating it. Although trained human observers have experience in spotting abnormalities or emergencies trivially, many physical and psychological factors act to reduce their performance. For instance, the CCTV cameras installed in a site always outnumber the observers employed sometimes by ten to one, therefore scattering their concentration and leading to a delay in detection or even misdetection. Furthermore, there is the additional factor of boredom due to the routine nature of the operation specially when significant incidents occur very seldom.

Many researchers in computer vision have considered the introduction of automation to existing CCTV systems as a solution to make the technology live up to public expectations. Consequently, automated surveillance systems have been conceived to either assist operators by alerting them to abnormal situations as they arise or replacing them in on-line pedestrian data collection systems. For an automated surveillance and crowd management system to be employed practically, it should offer an equal or better performance compared to traditional manual surveillance systems. Consequently, it is vital to perform extensive performance evaluations for such systems based on ground truths derived from human observer's performance.

---

* Department of Electronic Engineering, King's College London (University of London)

In this paper we give a brief presentation of an automated system to assist operators in underground stations to spot crowd-related emergencies. We then discuss the evaluation methods used to assess the system performance. We present three algorithms addressing the following situations:

- Detection of abnormal or forbidden direction of motion.
- Detection of suspicious stationary individuals or crowds.
- Crowd density estimation and overcrowding detection.

Section 2 introduces the system architecture and discusses the realisation of real-time processing. Section 3 describes the automated algorithms and their performance evaluation.

2 System architecture

The system consists of a Pentium 166 PC fitted with a black and white video digitiser (256 grey scale) and a motion detection board developed by the authors. Figure 1 shows the operation configuration.

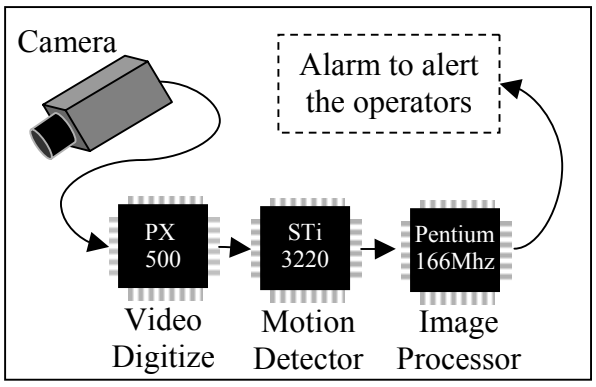


Figure 1: System architecture.

We realise real-time block-matching motion detection via a specialised hardware that operates on images of dimensions 512x512 pixels [10][11]. Image processing algorithms are applied to the

continuously updated motion vectors to perform feature extraction from motion. This approach benefits from:

- The reduction of the input data and consequently the processing cost by 128-fold whereby motion detection is performed with blocks of 8x8 pixels on pairs of video frames.
- The increase in the dimensionality of the input data to include speed and direction information.
- Reduced sensitivity towards variations in scene brightness.

The aspect of real-time processing is rather vital in this application, where prompt detection of emergencies is important to allow effective action to be taken. The processing rate for the system presented varies between 6 and 16 Hz depending on the complexity of the algorithm running. Due to the slow rate of change in scene events and the nature of the situations in interest, these figures are considered to be within the real-time processing requirement.

3 Performance evaluation of incident detection algorithms

Tests for system performance assessment are performed on live-camera and live-VTR (pre-recorded) video sequences. It should be noted that a VTR source normally has a worse signal-to-noise ratio (SNR) than a live camera (we have typically measured a reduction in SNR by 9dB at playback). Consequently, evaluation using a VTR provides a realistic worst-case scenario. Moreover, due to the indoor nature of underground stations, in some tests the scene brightness levels are altered synthetically (by varying the gain and offset of the video grabber) to test the stability of some of the algorithms under variations in scene brightness. The algorithms developed are integrated in a single demonstration system where the three main functionalities outlined above can be evaluated through a simple Graphical User Interface (GUI). Liverpool Street Station in the City of London was selected as the live verification site for the following reasons:

- It is situated in the city and has a busy railway link that makes it one of the major commuter paths to the city.
- The wide range of camera scenes available. There are 72 cameras installed, covering 70-80% of the station.
- The station has a spare control room that allows our live demonstration to take place without interfering with station control.

Figure 2: Operation room in Liverpool Street Station

Figure 3: Demonstration hardware installed on site

Figure 2 and Figure 3 show the operation room and the demonstration hardware installed on site during a live demonstration that took place on 6^{th} – 9^{th} of April 1998. Tests running on recorded video sequences are performed on the CROMATICA project video library at King's College London that includes more than 700 hours of recordings from different underground stations in London, Paris and Milan. The following sections present each of the three developed motion-based algorithms and the corresponding experimental performance evaluation.

3.1 Unusual or forbidden direction of motion

3.1.1 Algorithm

A simple polar classification of the motion vectors direction was initially considered for detecting the existence of forbidden trends of motion as in [1][3]. A background removal technique is

employed to eliminate noisy motion vectors prior to the classification process. This technique works well for scenes with good perspective view and horizontal motion paths. However, in the cases of bad perspective due to a low mounted camera (e.g. at the entrance or exit of a one-way corridor) the up-down oscillation that results from walking movements, is significant enough to give rise to false detection. Two solutions have been considered to this problem. First, the motion detection speed is reduced as an attempt to eliminate the effect of the oscillation by sampling at a low frequency. This approach did not yield good results because of the large variation in the up down oscillation frequency among individuals. Secondly, a region-growing algorithm is applied for grouping motion vectors according to their direction and to eliminate the small groups of vectors related to undesired movements. As this algorithm does not require any a priori knowledge of the scene background, it is also more suitable for scenes with varying brightness levels.

3.1.2 Performance evaluation

The region growing approach and the simple polar classification approach have been tested extensively on video sequences from a number of scenes that can be categorised as:

1. Ticket halls: main flow is sideways, good perspective view.
2. Corridors: up/down flows, bad perspective.
3. Ticket halls: varying brightness levels.

An evaluation measure is defined by the percentage of true, false and no detection events during a test's lifetime. An event is defined as an individual passing across the camera's field of view. Hence, the ground truth for assessment is obtained via a manual scanning of the test data by observers who are instructed with detection guidelines by experienced operators. Consequently, the terms **true, false** and **no detection rates** are defined as the number of individual incidents divided by the total number of events. The true, false and no detection rates correspond to true

positive, false positive and the true negative rates respectively. This assessment method is employed because it reflects the surveillance system aim to assess the behaviour of each individual passing through the interest area. Hence, an overall performance is integrated by assessing the system behaviour in each individual case.

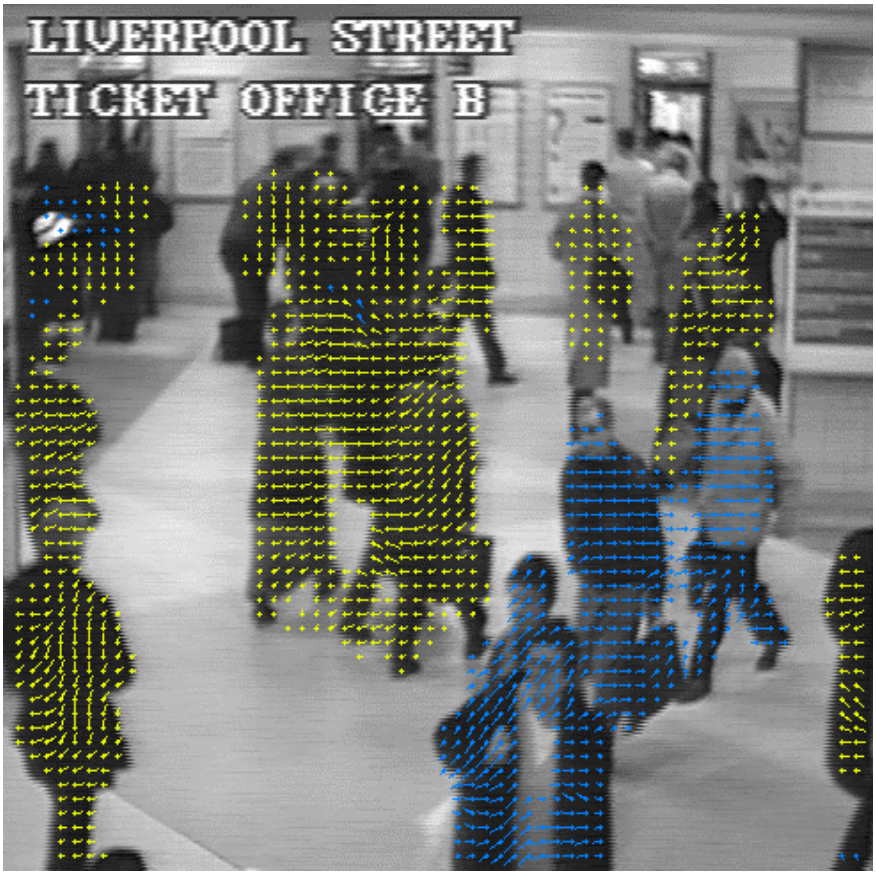


Figure 4: Motion direction detection in underground station's ticket hall with good perspective view.

For each of the two approaches tested, the algorithm parameters are tuned experimentally to yield the best results for the scene in question. Figure 4 shows the direction classification for individuals walking in a ticket hall with polar classification only. Figure 5 shows the classification of two opposing pedestrians walking along a corridor using region growing and polar classification. Table 1 shows the evaluation figures for the four tests performed.

3.1.3 Discussion

We conclude from the results of the tests shown in Table 1 that the region growing approach offers robustness in dealing with the up-down walking oscillation, unlike the simple polar classifier (although the latter shows a false alarm rate of 0% when working with good perspective views).

Approach	Ticket hall detection rate			Corridor detection rate		
	True	False	No	True	False	No
Polar classification only	99.12	0	0.88	100	11	0
Region growing + Polar	98.14	1.7	1.85	99.16	0.83	0.83

Table 1: Experimental evaluation figures for the algorithms of direction of motion detection.

Figure 5: Motion direction detection in a corridor with bad perspective.

Moreover, (synthetic) sudden variations in scene luminance have dramatic effects on both approaches. However, the region-growing algorithm can be made to ignore sudden brightness

changes efficiently by ignoring global motion variations unlike the former that requires rather expensive continuous background reference image updating.

Due to the indoor nature of the underground station's environment, brightness variations in the scene background are insignificant and have almost no effect on the performance of the algorithms.

3.2 Stationary individuals or crowds

3.2.1 Algorithm

Detection of stationary objects or persons in complex environments has been mainly addressed through three approaches, namely: temporal filtering [5], frequency domain methods [1] and motion estimation [3]. Our approach is related to what Bouchafa *et al* present in [3]. We present an algorithm that tackles the classic problems associated with stationary objects detection in complex scenes, namely:

- Frequent occlusion of the stationary object by moving pedestrians.
- Occlusion of the stationary object by moving pedestrians wearing colour shades similar to the background shades.
- Continuous change in pose and position of human subjects suspiciously waiting in public places.

A number of constraints are defined to detect the candidate stationary blocks and update their confidence at each iteration step. Moreover, variations in pose and position are updated within a few seconds (typically 3.25 sec.) by adopting a region growing technique. Accurate position updating is necessary for situations where there are stationary crowds to be detected yet few standing individuals are allowed. Hence, the detected stationary area is related to the corresponding number of people via a perspective distortion correction curve discussed in subsection 3.3.1 below.

3.2.2 Performance evaluation

Derived from the results of a survey carried-on within some underground stations in London, we define the normal period allowed for individuals to stand in underground stations' corridors or ticket halls (excluding queues) to be two minutes at the most. The algorithm is evaluated with different test data to verify its robustness against:

1. 100% occlusion. Complete occlusion by moving or standing pedestrians for unlimited periods of time. Forty-seven stationary pedestrians were examined.
2. Occlusion with the same colour as the background. Occlusion with moving pedestrians wearing grey shades similar to the background shades. Only eight cases were considered due to lack of data.
3. Pose and position variations. Movement of limbs and torso or shift in standing location with at least 1% overlapping with original position. Twelve cases were considered with an updating period of 3.25 seconds.

Moreover, the accuracy of the detection delay is assessed for each of the above tests. Consequently the evaluation metrics is defined as stability with occlusion, stability with occlusion with background shades, accuracy of detection delay and accuracy in updating the stationary area. In this evaluation process, an event is defined as the process of a pedestrian standing within the area of interest for more than the allowed period. Table 2 shows the performance figures for the tests mentioned above.

Test	Detection percentage		
	True	False	No
Normal Occlusion	97.9%	0%	2.1%
Occlusion with background colour	87.5%	0%	12.5%
Detection delay accuracy 2min ± 5sec	100%	0%	0%
Position updating in 3.25 Sec.	100%	4%	0%

Table 2: Performance figures for stationary object detection.

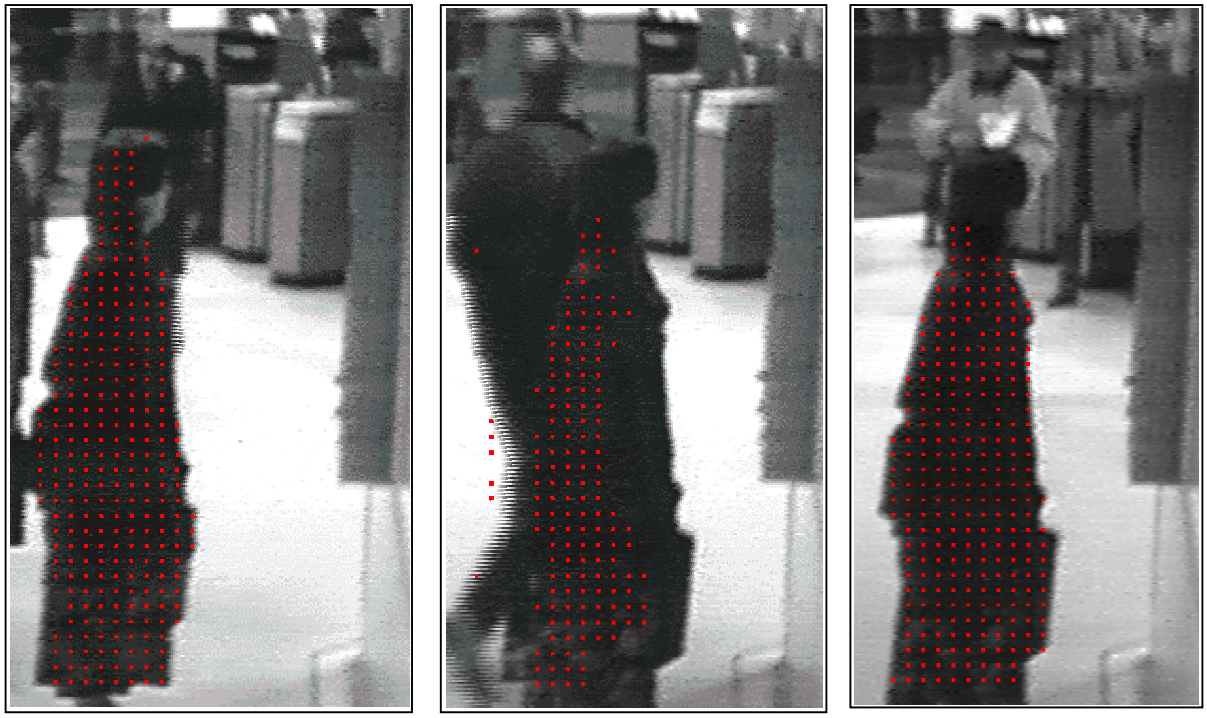
3.2.3 Discussion

The performance evaluation of the stationary object detection algorithm is obtained by examining the system's performance for each of the evaluation metrics stated above. This method is adopted to allow the tuning of algorithm parameters to maximise the performance of each module independently. Hence the use of the ground truth for the evaluation of this algorithm is test dependent, where the manual detection guidelines only state the maximum allowed stationary period.

Figure 6: Detection of occluded stationary pedestrians and updating their position.

From the figures of Table 2 we can reason that the system performs well with occlusion. The no detection rate is due to low contrast with the background image rather than occlusion. On the other hand, the stability for occlusion by moving pedestrians wearing grey shades similar to the background shades is less impressive, because pedestrians walking slowly are confused with the stationary background. However, only one event out of the eight examined was erroneous. The detection delay accuracy is adjusted by taking the processing time into consideration, therefore overlooking the processor time. Nevertheless, the detection delay period of two minutes is

achieved with a tolerance of ± 5 seconds due to errors introduced by pedestrian position variations. Finally, a few false detection cases were experienced due to the position-updating process. Hence, occluded pedestrians will pass their stationary period count to the ones in the background as soon as they move, causing a shorter detection delay.



(a) Pedestrian detected after standing for thirty seconds.

(b) Pedestrian moves closer to see a sign after twelve seconds.

(c) Position updated after approximately three seconds.

Figure 7: Pose and position updating sequence.

In Figure 6 a couple of pedestrians standing to chat for more than two minutes are detected despite continuous occlusion by passing crowds. On the other hand, the bag held by one of the women is not detected due to the small contrast with the background. Figure 7 shows the performance of the position updating process in (c) after a change in position from the initial detection in (a). The position updating process does not account for any changes in pedestrian size. That is, it attempts to detect the variations in position by assuming fixed occupied area. Therefore, the head position in (c) is not recovered properly after a movement towards the camera

that increases the occupied area. The fixed area constraint is necessary to prevent the region growing process from covering nearby pedestrians.

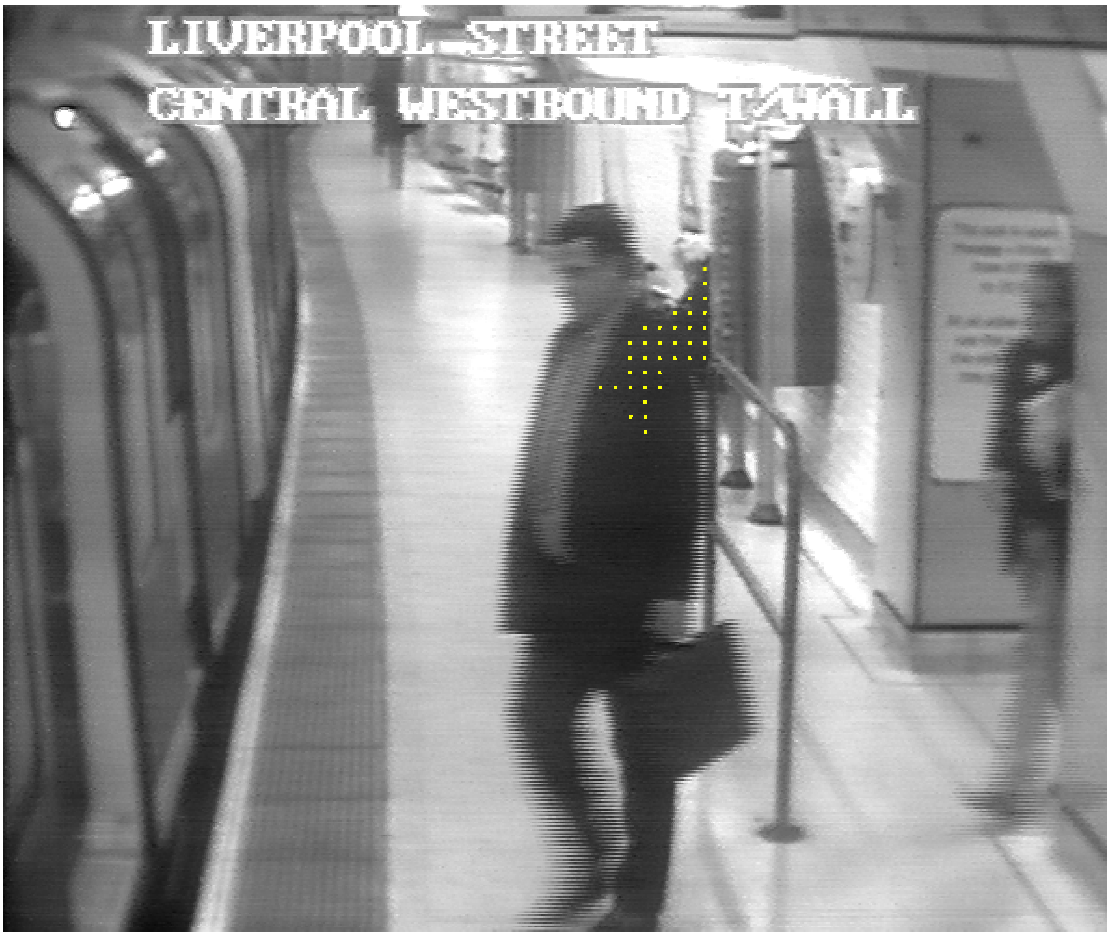


Figure 8: Detection of a passenger sitting in a platform despite frequent occlusions.

Detection of occluded stationary passengers is shown in Figure 8, where the passenger detected remained sitting in the platform for more than the set period.

3.3 Crowd density estimation

3.3.1 Algorithm

Automated estimation of crowding levels in public places has gained a lot of interest [1][2][4][6][7][8][9][12][13][14] because it plays a big role in ensuring public safety. One of the main ideas in the field of computer vision to estimate crowding levels (or to count people), is to establish a direct relationship between the number of feasible image features (e.g. edge pixels,

vertical edges, foreground pixels, circles, blobs etc.) and the crowding level (or the number of people in the scene). Here we follow the steps of Velastin *et al* in [14] and Tsuchikawa *et al* in [12] by using a background subtraction technique to identify the foreground image blocks as features to estimate the number of people in the scene. However, because the global measurement procedure employed assumes a homogeneous pedestrian distribution, crowd density estimates given via this approach are erroneous, as perspective distortion is not taken into consideration.

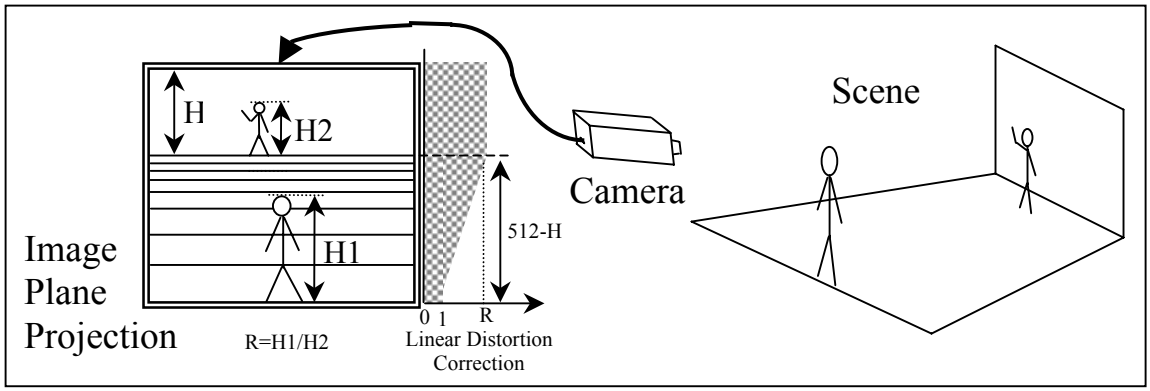


Figure 9: Imaging system and image plane projection for perspective distortion correction.

To compensate for this effect, an appropriate weighting procedure is required to scale the number of detected features by an inverse perspective effect. Hence, we propose a non-linear curve derived from on-line camera calibration to modify the count of foreground image blocks according to their vertical position in the image. Figure 9 defines the imaging system and the inverse perspective distortion introduced. Moreover, it shows a linear correction curve corresponding to the image plane. On the other hand, Figure 10 shows the employed non-linear perspective correction curve derived from the above linear curve through camera calibration. A second approach is adopted to estimate crowd density via counting stationary people with a stationary detection delay of thirty seconds. This idea is derived from the observation that "overcrowding is inversely related to the pedestrians' flow speed".

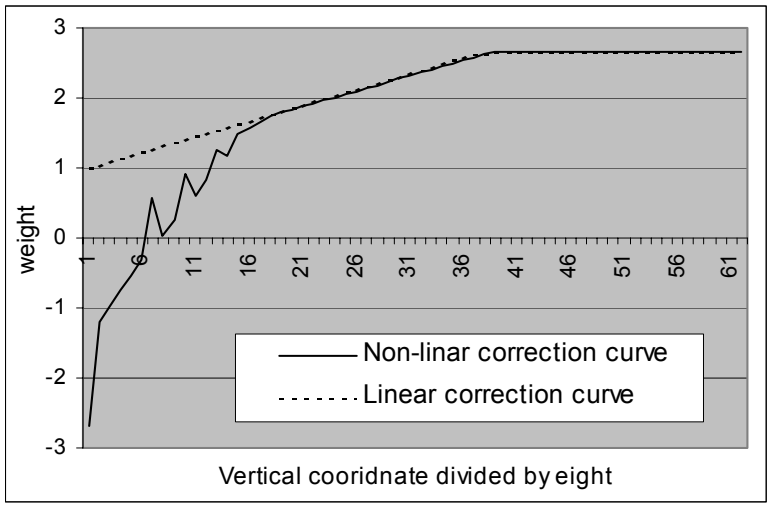


Figure 10: The linear and the updated non-linear perspective correction curves for ticket hall B in Liverpool Street Station London.

Hence, a period of thirty seconds is selected as a good compromise to satisfy the trade-off between detection delay and false detection rates for the particular scene under study. Experimental limits for the maximum number of people in the scene before overcrowding occurs, have been obtained and employed as thresholds for sounding the alarms.

3.3.2 Performance evaluation

The evaluation begins by verifying the goodness of the perspective-distortion correction process. Therefore, the curve is tested for all possible crowding levels. Figure 11 shows the system's response compared to the manual counts. The test data for the evaluation of the crowd density estimation algorithms is a video sequence that lasts for three hours and includes most, if not all, possible crowding levels expected in the scene. A manually generated log file that includes the number of people in the scene at fixed time intervals (every three seconds) was created as a ground truth for performance evaluation of the two algorithms presented.

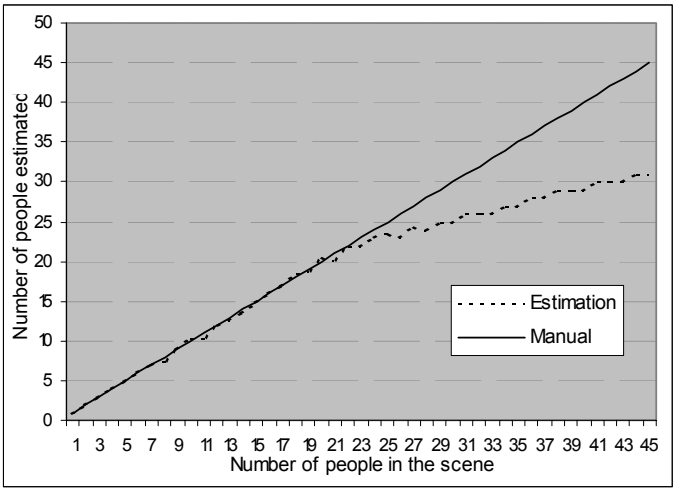


Figure 11: Estimated number of people for the automated and manual detection.

Consequently, a threshold is defined, by following experienced operators' guidelines to identify the overcrowded situations, hence comparing them with the automatic detection's alarms. Figure 12 shows the manual estimation for the number of people in the scene during the last hour of the test together with the automatic estimations generated by each of the two algorithms.

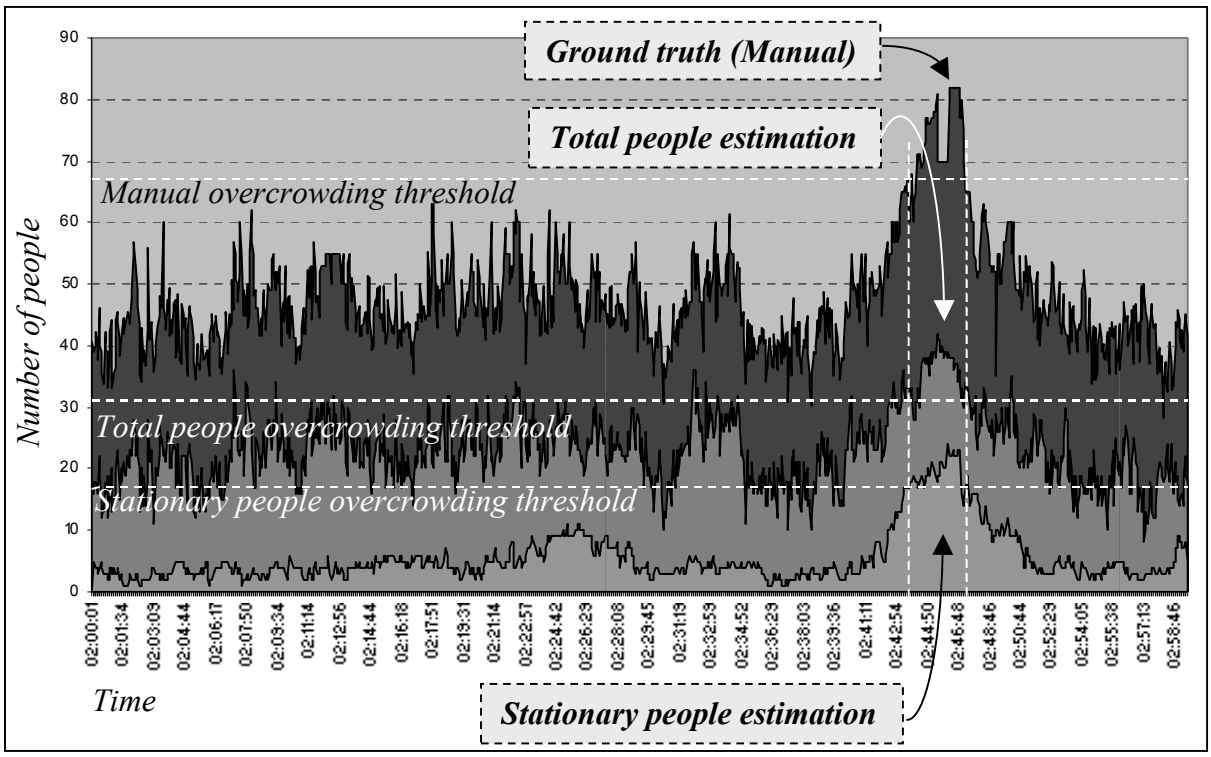


Figure 12: Manual and automated crowding levels estimation. Overcrowding limits are defined via horizontal dotted lines, and overcrowded situations are enclosed by vertical dotted lines.
(The ground truth curve is shifted upwards by twenty units for clarity)

To estimate the first algorithm's performance (estimating the total number of people in the scene) as a means for counting people, the automatic detection figures are directly compared with the manual ground truth leading to very weak performance at high crowding levels due to the severe effects of occlusion. However, the objective of the evaluation procedure should consider the assessment of the algorithms' performance towards the detection of overcrowded situations as the ultimate aim of the system. Consequently, the situations classified by the automated algorithms as being overcrowded are compared with the manual classification and the true, false and no detection rates are calculated. An event is defined as the process of classifying a video frame and generating an estimate of the crowding level in the scene. 1200 events were experienced during the three hours test, and the performance figures for the algorithms are presented in Table 3.

Approach	Overcrowding detection		
	True	False	No
Estimating the total number of people	95.62%	4%	0.38%
Estimating the number of Stationary people	98.51%	0.28%	1.21%

Table 3: Performance figures for automated overcrowding detection.

A **true** detection is defined as the situation where the automated algorithm classification matches the manual classification, while **False** detection is referred to cases classified as crowded by the automatic system and not by the manual observation, and vice versa for the **No** detection cases.

3.3.3 Discussion

The poor performance reflected by the perspective distortion correction curve at high crowding levels is due to occlusion. This is inherent in the adopted approach, where the features selected to estimate crowd density (foreground image blocks) are not immune to occlusion effects. On the other hand, the performance of the updated curve is tested against the linear curve, showing that the later has a poor performance when pedestrians are close to the camera whereas the former gives accurate results.

The non-linear relationship between the estimated number of people and the actual figures as in Figure 11 shows a decline in estimation accuracy as crowding levels rise. Hence, the actual process of overcrowding detection is operating at the non-linear stretch of the response because the decision making margin lies at high crowding levels. However, the deterioration of the performance figures for the automated classifiers, as shown in Table 3, is not owed to the non-linearity in the correction curve response, as much as to the behaviour of the individual algorithms towards different crowding situations. The first algorithm estimates the total number of people in the scene at sixteen times per second; consequently any rise in the area occupied by pedestrians is noticed immediately and translated to a rise in crowding levels. Hence, false alarms are sounded when a moving crowd that is loosely distributed covers most of the area under observation. Therefore, the false alarm rates are high, and exceed the desired rates by the end users, whereas the no detection rate is very low due to the fast system response. The second algorithm estimates the number of stationary people with a stationary detection delay of thirty seconds; consequently fast variations in pedestrians density are ignored and the slow variations are translated to measure crowding levels. Therefore, the false detection rates are very low, whereas the no detection rates increase due to the slow system response.

Large moving crowd
(false overcrowding detection).

Overcrowded ticket hall
(true detection)

Figure 13: The classification of the crowding levels by estimating the total number of people in the scene.

Figure 13 and Figure 14 show some examples of true and false detection cases by the evaluated algorithms. The bright dots on the images correspond to image blocks employed in crowding levels detection. Consequently, the two algorithms' performance can be compared from the figures in Table 3, yielding the conclusion that counting stationary people in the scene will offer more stability and robustness although it will experience a small delay in detection.

Large moving crowd not detected as an overcrowded case.

An overcrowded scene that is detected with some delay.

Figure 14: Classification of the crowding levels by estimating the number of stationary people in the scene.

4 Conclusions

In this paper we presented experimental evaluation methods and results for three motion-based algorithms for the detection of emergencies in crowded conditions. Three distinct evaluation methods are adopted. First, the performance of two algorithms for the detection of forbidden direction of motion is compared and tested under different imagery and scene conditions. Secondly, the performance of the stationary people detection algorithm is tested for four individual factors as a method to evaluate the overall performance. Finally, two algorithms for overcrowding detection are compared based on direct comparison with the manually generated ground truths.

5 Acknowledgement

The work described in this paper was carried out as part of the EC project TR-1016 "CROMATICA". Partners in this project are UCL, RATP, LUL, Politecnico di Milano, ATM, Molynx, INRETS, USTL and CEA-LETI.

6 References

[1] C. Davies, J.H Yin, S. A. Velastin 'Crowd Monitoring Using Image Processing' Electronics & Communication Engineering Journal, February 1995.

[2] S. A. Velastin, A. C. Davies, J.H Yin, M. A. Vicencio-silva, R. E. Allsop, A. Penn 'Analysis of crowd movements and densities in built-up environments using image processing' IEE Colloquium on Image Processing for Transport Applications 1993 pp.8/1 - 8/6.

[3] S. Bouchafa, D. Aubert, S. Bouzar 'Crowd Motion Estimation and Motionless Detection in Subway Corridors by image processing' IEEE Conference on Intelligent Transportation Systems 1997 (ITSC'97) pp. 332-337.

[4] S. Regazzoni, A. Tesei, V. Murino 'A real-time vision system for crowding monitoring' Proceedings of the International Conference on Industrial Electronics 1993 (IECON'93) pp/ 1860-1964 vol.3

[5] M. Takatoo, C. Onuma, Y. Kobayashi 'Detection of objects including persons using image processing' proceedings of the 13th IEEE international conference on Pattern Recognition pp.466-472 vol3.

[6] A.J. Schofield, T.J. Stonham, P.A. Mehta 'A RAM based neural network approach to people counting' Image Processing and Its Applications, 4-6 July 1995. Conference Publication No.410 © IEE 1995.

[7] T. Coianiz, M. Boninsegna, B. Caprile. 'A Fuzzy Classifier for Visual Crowding Estimates', IEEE International Conference on Neural Networks, 1996. pp. 1174 - 1178 vol.2 , 3-6 June 1996

[8] Ottonello, M. Peri, C. Regazzoni, A. Tesei 'Integration of Multisensor DATA for Overcrowding Estimation' 1992 IEEE international conference on systems, man, and cybernetics, vols 1 & 2, 1992, ch. 309, pp.791-796

[9] C. Regazzoni, A. Tesei 'Density Evaluation and tracking of Multiple Objects from Image Sequences' in Proceedings of IEEE International Conference on Image Processing, 1994 (ICIP-94) Volume: 1 , Pages: 545 -549 vol.1

[10] B. A. Boghossian, S. A. Velastin 'Real-time motion detection of crowds in video signals' IEE Colloquium on High Performance Architecture for real-time image processing, February 1998 p.p 12/1- 12/6.

[11] B. A. Boghossian, 'Real-time motion detection in video signals' MSc thesis, department of Electronic Engineering King's College London October 1997.

[12] M. Tsuchikawa, A. Sato, H. Koike, A. Tomono 'A Moving-object extraction method robust against illumination level changes for pedestrian counting system' Computer vision 1995 Proceedings of the International Symposium on pp.563-568.

[13] N. Marana, S. A. Velastin, L. F. Costa, R. A. Lotufo 'Estimation of crowd density using image processing' Image Processing for security applications 1997 IEE Colloquium on pp 11/1 – 11/8.

[14] S. A. Velastin, J. H. Yin, A. C. Davies, M. A. Vicencio-Silva, R. E. Allsop, A. Penn 'Automated measurement of crowd density and motion using image processing' Road traffic monitoring and control 1994, Seventh International conference on, pp.127 – 132.