Spatiotemporally Localized New Event Detection in Crowds

Alexia Briassouli, Ioannis Kompatsiaris
Informatics and Telematics Institute
6th km Charilaou Thermis,
Thermi 57001, Thessaloniki, Greece
abria@iti.gr,ikom@iti.gr

Abstract

The behavior of crowds is of interest in many applications, but difficult to analyze due to the complexity of the activities taking place, the number of people moving in the scene and occlusions occurring between them. This work focuses on the problem of detecting new events in crowds using an original approach that is based on properties of the data in the Fourier domain, which leads to computationally effective and fast solutions that lead to accurate results without requiring data modeling or extensive training. The PETS2009 dataset has been used for benchmarking algorithms developed for analyzing crowd behavior, such as recognizing events in them. Experiments on the PETS 2009 dataset show that the proposed approach achieves the same or better results than existing techniques in detecting new events, while requiring almost no training samples. Extensions for accurate recognition and dealing with more complex events are also proposed as areas of future research.

1. Introduction

Videos with crowds of people walking are difficult to analyze using traditional methods due to the complexity of the motions present, occlusions, and changes in the types of activities occurring. Recently, this problem has attracted research attention as the recognition of events in crowds can be very useful for surveillance and monitoring, e.g. for ensuring safety in public places or for evacuation planning [10]. The PETS 2009 dataset has been used in many works for benchmarking purposes (also in PETS2010) and will be used here to evaluate the results of the proposed method.

Videos of crowds are considered to fall in the category of temporal textures, i.e. textures which evolve over time, and whose motion is difficult to analyze due to its highly non-rigid nature. Several approaches have been developed to analyze crowd behavior in the PETS 2009 dataset, where a wide range of events occurs. Some research focuses on people counting and tracking, while others, including this work, detect events in crowds. Methods that focus on the behavior of the individuals in the crowd are known as microscopic, such as the Social Force model [13], [17] and others [5]. Several approaches search for deviations from a path a person should follow [8], [9], [20], however they require scene segmentation and tracking, which can be very difficult and costly in the crowd motion environment. Additionally, they make the assumption that the crowd (or other moving entity) has a fixed path or trajectory, which limits their general applicability.

Macroscopic methods that treat video frames and the crowds moving in them as a whole are more appropriate for examining group behavior and can model the scene of interest with lower complexity [11], [6]. In [15], [3], Hidden Markov Models (HMMs) are built to separate normal from abnormal motions. The drawback of HMM-based approaches is the requirement that the optical flow should be characterized by conditional independence, which is not realistic, particularly in crowded environments. Conditional Random Fields (CRFs) are used instead in [19] for separating normal from abnormal behavior, without the requirement of conditional independence of the flow. Lagrangian particle dynamics have also been used for crowd flow segmentation and detection of instabilities in crowd flow [2]. This approach requires the estimation of the flow field and examines the evolution of particle trajectories, so it examines the data both at a microscopic and macroscopic level. In [21], [22] low-level information is used for event modeling and behavior understanding in video, without requiring tracking. These methods are complex and therefore have an offline nature, unlike the approach proposed in this work.

In this work, an original method for the detection of changes in crowd behavior is presented, which is not based on generative modeling of the data, does not require the prior estimation of the optical flow and which has limited training requirements. The model of the motion in the scene is derived from the Fourier transform (FT) of the data, on

2011 IEEE International Conference on Computer Vision Workshops
978-1-4673-0063-6/11/$26.00 ©2011 IEEE

which statistical sequential change detection methods are applied to detect new events taking place. The phase of the FT has been used successfully for dynamic texture modeling and particularly for recognition and synthesis in [12]. The phase information of the FT is used here as well, but with a different theoretical basis, namely in order to produce an accurate, non-parametric statistical model for the random motions taking place in the scene. Also, the purpose of this work is to detect new crowd behaviors, unlike the work of [12], which focuses on modeling, recognition and synthesis.

The motion in crowds is considered to follow a random distribution based on properties of the FT, as detailed in Sec. 2, without requiring the estimation of the optical flow in the scene, nor any parametric modeling or model fitting. The resulting non-parametric statistical model offers a complete description of the random motion characteristics for a temporal texture, in this case a moving crowd. Consequently, it can be used to detect when a change takes place in the video from one temporal texture to another, and in practice from one type of crowd motion to another. This is achieved by using statistical sequential change detection techniques, as described in Sec. 3. Sequential change detection methods are designed precisely for the problem of detecting the unknown moment of change from one distribution to the next, by using each sample (video frame) as it arrives. Thus, a real-time approach for the detection of new events in crowd motions is developed, with minimal training requirements and accurate detection results, as Sec. 5 shows.

This paper is organized as follows. In Sec. 2, a mathematical model of a video of temporal textures (like moving crowds) in the Fourier domain is presented. Sequential change detection methods for finding moments of change between temporal texture subsequences is presented in Sec. 3. The application of this method on a global and local scale, depending on the needs of the problem, and an overview of the proposed method, are presented in Sec. 4. Experimental results for videos with various crowd motions from the PETS2009 dataset are presented in Sec. 5. The crowd is considered at both a global and a local level, to examine the performance of the proposed approach in each case. Conclusions and future work are presented in Sec. 6.

2. Fourier Domain Random Motion Modeling

The motivation for basing our method on the Fourier domain is the observation that motion information is concentrated in the Fourier phase [12], [14]. The motion of crowds, and temporal textures in general, can be considered to follow an unknown random distribution. It will be shown below that a complete description of the random motion characteristics can be extracted from the Fourier domain without resorting to optical flow estimation. Consider a video with illumination $c(x, y)$ at frame 1 in the spatial domain at pixel (x, y), which undergoes random displacements $r_x(t), r_y(t)$ following the distributions f_x, f_y respectively. For simplicity of notation, we will omit (t) from the displacements, which also emphasizes that they change due to their random nature rather than as a function of time. Frame t in the spatial domain can then be expressed as:

$$c(x, y, t) = c(x, y)\delta(x - r_x t, y - r_y t) \tag{1}$$

The FT in all three dimensions, namely space (x, y) and time (t) is estimated [14] by:

$$C(u, v, \omega) = C(u, v)\delta(\omega + ur_x + vr_y), \tag{2}$$

where $C(u, v)$ is the two-dimensional spatial FT of the first frame. From the inverse one-dimensional FT in time, we can obtain $C(u, v, t)$ as follows:

$$\begin{aligned} C(u, v, t) &= \int C(u, v, \omega)e^{j\omega t}d\omega \\ &= \int C(u, v)\delta(\omega + ur_x + vr_y)e^{j\omega t}d\omega \\ &= C(u, v)\int \delta(\omega + ur_x + vr_y)e^{j\omega t}d\omega \\ &= C(u, v)e^{-jur_x}e^{-jvr_y}, \end{aligned} \tag{3}$$

where r_x, r_y are random displacements at time t that characterize the crowd motion (or temporal texture in general). $C(u, v, t)$ is the 2D spatial FT of video frame t and is a function of the random displacements r_x, r_y that follow the pdf's f_x, f_y, respectively. From Eq. (3), we get:

$$L(u, v, t) = \frac{C(u, v, t)}{C(u, v)} = e^{-jur_x}e^{-jvr_y}. \tag{4}$$

Assuming there are many instantiations of the same crowd video providing many samples of the randomly displaced video frames $c(x, y, t)$, one could obtain several instantiations of $L(u, v, t)$. If the arithmetic average is considered to approximate the ensemble average $E[\cdot]$, and if the x and y displacements are assumed to be independent (i.e. moving right/left is independent of moving up/down), we have:

$$E[L(u, v)] = E[e^{-jur_x}e^{-jvr_y}] = E[e^{-jur_x}]E[e^{-jvr_y}], \tag{5}$$

where $E[e^{-jur_x}] = \int f_x(r_x)e^{-jur_x}dr_x$, $E[e^{-jur_y}] = \int f_y(r_y)e^{-jur_y}dr_y$. From probability theory, it is well known that the characteristic function of a random variable y with pdf $f_Y(y)$ is given by its complex conjugate FT:

$$\Phi_Y(\omega) = \mathfrak{F}^*[f_Y(y)] = \int f_Y(y)e^{j\omega y}dy = E[e^{j\omega y}]. \tag{6}$$

The pdf of y can then be estimated from the characteristic function via the inverse FT as follows:

$$f_Y(y) = \frac{1}{2\pi}\int \Phi_Y^*(\omega)e^{-j\omega y}d\omega. \tag{7}$$

Then, if the characteristic functions of the distributions for r_x and r_y are $\Phi_{r_x}(u)$ and $\Phi_{r_y}(v)$, Eq. (5) becomes:

$$E[L(u,v)] = \Phi_{r_x}(u)\Phi_{r_y}(v). \quad (8)$$

The characteristic functions of the random motions appearing in the video can then be estimated from the data FTs without estimating and modeling the actual optical flow. This is particularly useful, as the characteristic function provides the most complete statistical description of a random variable since its pdf and all existing moments can be extracted from the characteristic function. If it is necessary to approximate the pdf of the motion only in the x or only in the y direction, one simply estimates $E[L(u,0)]$ or $E[L(0,v)]$ respectively, since $\Phi_Y(0) = \int f_Y(y)dy = 1$.

2.1. Characteristic Function Estimation

In practice, when analyzing a video of crowd motions, there are not many instantiations available for the approximation of the characteristic function. In order to overcome this, the realistic assumption is made that the motion statistics do not change significantly within a window of w_0 video frames around a time instant t of interest. Thus, in practice we use w_0 video frames around frame t to approximate the characteristic function:

$$E[L(u,v)] = E\left[\frac{C(u,v,t)}{C(u,v)}\right] = \frac{E[C(u,v,t)]}{C(u,v)}, \quad (9)$$

where

$$E[C(u,v,t)]) = \sum_{k=t-w_0}^{t} C(u,v,k). \quad (10)$$

Then, the probability density function of the motion in the x and y directions can be estimated from the inverse FT of the approximated characteristic function as in Eq. (7). Our experimental results showed that this approximation provides good modeling results, as changes are detected with accuracy in the crowd motions (see Sec. 5).

3. Sequential Change Detection

The motion of a temporal texture, e.g. a crowd, contains the most essential information about the type of activity taking place: for example, a crowd may be walking, running, dispersing etc. The change from one type of motion to another is of practical interest for surveillance and monitoring purposes. Statistical sequential change detection such as the Cumulative Sum (CUSUM) method is well suited for detecting such changes, since it is able to detect a change in a dataset's pdf in real time, i.e. as each new data sample arrives. If the data under examination follows a different kind of motion after an unknown moment of change t^*, the likelihood ratio at time k is estimated by $L_{1,k} = \frac{f_{1,k}}{f_0}$, where f_0 is the initial data pdf and $f_{1,k}$ is the approximation of the data pdf at the current instant k. In our case f_0 is approximated via the characteristic function (see previous section) over the frames of the data-set that undergo the "baseline" (motion, and a change is found when the crowd motion deviates from it. The likelihood ratio is used for the test statistic of the CUSUM test at frame k, expressed in the computationally efficient iterative form [18]:

$$T_k = \max(0, T_{k-1} + L_{1,k}), \quad (11)$$

where $T_k = T_{1,k}$ and $T_0 = 0$. When the current data pdf deviates from f_0, the values of test statistic T_k increase significantly, and a change can be detected at that point.

In practice, the threshold indicating that the change in T_k corresponds to an actual change is estimated from training data, with the goal of leading to the accurate detection of changes with few false alarms. For the videos of PETS2009 we obtained good results for the threshold $\eta = \mu_k + c \cdot \sigma_k$, with $c = 1.5$, where μ_k, σ_k are the mean and standard deviation of the test statistic values until frame k. The estimation of the test statistic and changes in the crowd motion uses only current data values, making the proposed method appropriate for real-time implementations.

4. Spatially Global or Local Motion Change Detection

The frequency-based modeling of Sec. 2 produces a pdf of the motion characterizing the video frame upon which the FT is applied. Thus, application of the change detection method presented in Sec. 3 will lead to the detection of changes in the overall frame motion. In the experiments of Sec. 5.2 we applied our approach to entire video frames in order to detect changes in the crowd motions holistically, and produce results that are comparable with, or outperform, those of [6], [17] and [19]. However, applying frequency domain methods to the entire frame results in loss of spatial localization, as they result in a pdf that characterizes the entire frame motion and cannot deal with multiple temporal textures. In many applications it is of interest to detect changes in different parts of each video frame, which may contain, e.g., different groups of people whose motions undergo various changes. In order to overcome the issue of spatial localization, we apply the method of Sec. 2 in a block-wise manner over each frame, as in [1], [7], [19]. Then (1) The FT is calculated over 8×8 blocks to find the block-wise pdf as in Sec. 2. (2) Likelihood ratios are estimated by blockwise comparison of the data pdf with that of the training data. (3) The block-wise likelihood ratios form the CUSUM test statistic T_k (Eq. (11)) for frame k. (4) T_k is compared with the threshold η to determine if a change has occurred. Blocks of size 8×8 are chosen as they provide enough data for the estimation of the motion pdf, while at

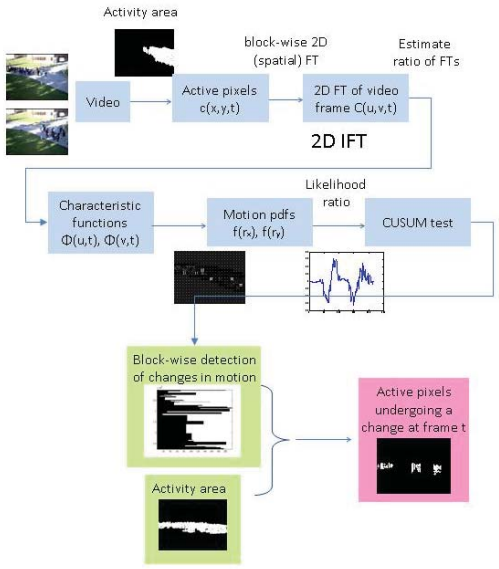


Figure 1. Block diagram of spatially localized FT-based change detection for motions of temporal textures.

the same time remaining limited in their spatial extent and maintaining spatial localization.

This procedure detects the moments of change for each block in each frame, leading to a binary "change mask" per frame, whose value is one at a block that has undergone a change. In order to remove possible false alarms, a binary mask showing which pixels are active is also estimated for $w_0 = 20$ frames before t. This mask can be extracted using any background removal method, such as those presented in [16]. In this work the kurtosis-based method of [4] is used to localize active pixels, due to its simplicity and robustness. The change mask and the activity mask are combined using the "and" operation, providing a binary mask for each frame with values equal to one at active pixels that undergo a change in their activity. Thus for each video frame a binary mask of pixels undergoing a change is extracted, providing spatiotemporal segmentation that localizes new events in new locations. The video of temporal textures, or crowd motions, is then separated into regions where changes in motion are detected and whose motion can be characterized before and after each change. This procedure is depicted in the block diagram of Fig. 1.

5. Experiments

Experiments take place with videos of PETS2009 containing crowds that undergo various motions and changes in them. We examine the videos that have been used in [6], in order to compare the accuracy of change detection results. In each experiment, the video is processed both as a whole and in a block-wise manner. In the first case, moments of change in the overall crowd motion are found. Videos of the results can be found at $http://mklab.iti.gr/mklab\_people/\ abria/CrowdsEvents/demos.html$.

5.1. Ground Truth

A ground truth for the sequences of PETS2009 is provided in [6] for walking, running, evacuation, merging and other events. However some transitions are gradual, as also noted in [19], and our ground truth for them does not always coincide with that of [6]. For example, in [6], for the $S3 : 14 - 16$ sequences, they claim that there is walking from frames $0 - 30$ and $108 - 162$ and running in the others. Actually, very few people from the front of the crowd start to run at frame 37, as the bulk of the crowd (more than half of it) starts to run after frames $50 - 60$. Additionally, at frame 162 only the first one or two people start to run; we consider that a change to running takes place at frame 180, when more than half the crowd is running. Also, the ground truth in [6] does not always correspond to all views. For example, in $S3 : 14 - 33$, two walking people that join the crowd are visible after frame 240 in view 2, but after frame 280 in view 4. These frames are all shown in Fig. 2. In the sequel, we consider as ground truth (G.T.) for $S3 : 14 - 16$ the frames $50, 108, 180$, for $S3 : 14 - 31$ frames $50, 130$ and for for $S3 : 14 - 33$ the frames $265, 330$.

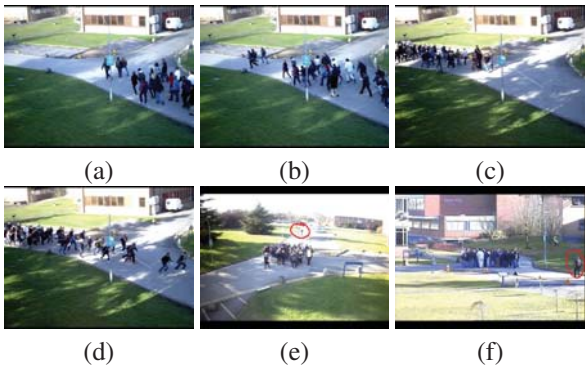


Figure 2. Video frames. $S3 : 14 - 16$, view1: (a) 37 , (b) 50, (c) 162, (d) 180. (e) $S3 : 14 - 33$, view2. People enter the scene at frame 240 in the red circle. (f) $S3 : 14 - 33$, view4. People enter the scene at frame 280 in the red circle.

5.2. New event detection over entire crowd

The videos S3 from PETS2009 contain a group of people walking, then running and exiting, a group of people merging and then dispersing, i.e. they contain crowds of people whose motion undergoes many changes. An important issue is the clear definition of the "normal" or baseline motion, from which a change occurs in each case. In this set of videos, we consider as baseline motion that of the first 20 video frames of the video. When a change is detected at a

Sequence S3	Det. Changes	Our ground truth
14-16 view 1	52, 101, 156, 192	50, 108, 180
14-16 view 2	63, 98, 141, 167	”
14-16 view 3	42, 69, 160	”
14-31 view 1	38, 129	50, 130
14-31 view 2	77, 125	”
14-31 view 3	71, 96, 123	”
14-31 view 4	55, 60, 125	”
14-33 view 1	164, 284, 348	265, 330
14-33 view 2	100, 172, 260, 337	”
14-33 view 3	171, 231, 319	”
14-33 view 4	121, 284, 328	”

Table 1. New event detection in PETS2009. Discrepancies are due to the gradual nature of the transitions.

Methods	Results (%)
our method	96.4
Pathan [19]	97.1
Mehran et al. [17]	96
Chan et al. [6]	81

Table 2. Percentage of correctly detected new events.

frame t_{ch}, the baseline motion is *reset* to be that of the first 20 frames after t_{ch}.

In this section new events are detected over the entire frame, so the crowd motion is treated as a whole. Table 1 shows changes that are detected in our work and the ground truth. In some cases the detected changes do not coincide numerically with the ground truth because of the gradual nature of the transitions, where the ground truth moment of change is somewhat subjective. The videos in our demo site show that the detected changes indeed correspond to a change in motion. Table 2 shows a comparison of the detection accuracy for our method and the methods of [6], [17], [19], where it can be seen that it performs nearly as well or better than existing techniques. It should be noted that, unlike [19], our method has minimal training requirements (only 20 frames after the most recent change are used as baseline for comparison). Also, for $S3:14-16$, view 2, a change is also detected at frame 141. This is not contained in the ground truth, but at that frame a person departs from the main group and walks to the left, as seen in Fig. 3(a) (in the red circle). In Fig. 3 we show characteristic frames of the changes in the crowd motion that are detected and shown in Table 1.

5.3. Spatiotemporally Localized Changes

As discussed in Sec. 4, spatiotemporally localized changes, i.e. new events, can be detected when the method of Sec. 2 and 3 is applied in a blockwise manner to

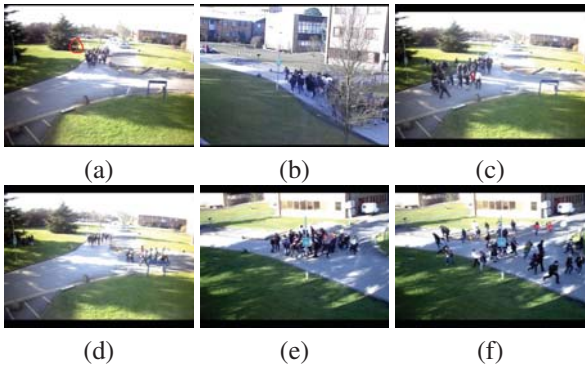


Figure 3. Video frames. $S3:14-16$, view 2: (a) 141. Video $S3:14-16$, view 3: (b) 42. $S3:14-31$, view 2: (c) 77, (d) 125. $S3:14-33$, view 1: (e) 260, (d) 348.

the video. Videos showing the spatiotemporally *localized* changes, i.e. pixel regions which undergo changes at each time instant can be seen in the supplementary material (the videos in the folder “ST localized changes”). There are changes in a pixel region when a transition occurs e.g. from no motion to walking, then running or, again, no motion, and so on.

For computational purposes, the blockwise likelihood ratio estimated for each $N_1 \times N_2$ video frame is vectorized, so that the video is transformed into a two-dimensional $N_1 \cdot N_2 \times N$ image (for N frames), where the variation of the likelihood ratio over time can be seen clearly. The threshold η of Sec. 3 is then applied to it, producing a binary mask indicating at which time each pixel undergoes a change. An example of the resulting mask is shown in Fig. 5(a). This mask is then transformed back into a two-dimensional image for each time instant, creating a three-dimensional binary “change mask” for the video. As described in Sec. 4, and also shown in Fig. 1, this mask is combined with the activity mask on each video frame, so as to isolate the moving pixels which have undergone a change, as in Fig. 5(b), (c) and also in the supplementary material. In this manner, we detect which active regions in each video frame undergo a change in their motion, and can detect spatially local new events, although a frequency-domain method is being used.

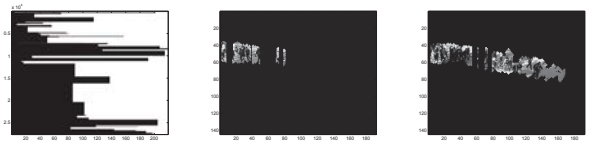
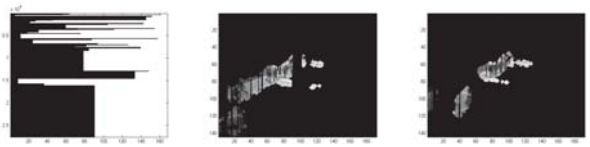


Figure 4. Video $S3:14-16$, view 1: walking, running. (a) Binary mask of changes per pixel. The horizontal axis shows time and the vertical axis shows the pixels. (b), (c) Frames 80, 180 with the pixels that have undergone a change.



(a) (b) (c)

Figure 5. Video $S3 : 14 - 46MV$, view 2: walking, running. (a) Binary mask of changes per pixel. The horizontal axis shows time and the vertical axis shows the pixels. (b), (c) Frames 125, 140 with the pixels that have undergone a change.

6. Conclusions

In this work an original approach to the problem of new event detection in videos of crowds is presented. The proposed approach uses the Fourier transform to model the random crowd motion in a manner that does not require estimation of the optical flow, parametric modeling, or extensive training. Real-time detection of changes in the motion of the crowds' motions is achieved through the application of statistical sequential change detection techniques, and in particular the CUSUM method. This leads to the accurate detection of changes in the crowd motions at a global, scene-wise, level. Spatiotemporal localization of the changes in crowd motion is also achieved through the block-wise application of the proposed method. Experiments show that accurate results are attained in both the global and local problems, with accuracy as good or better than that of existing methods. Future areas of work include the utilization of the derived motion probability distributions for activity recognition, which can lead to the spatiotemporal labeling of pixel areas in the video as the crowd motion in them evolves. Also, simple extensions, involving further processing for the minimization of false alarms and the elimination of the effect of shadows, will take place to refine the spatiotemporal segmentation results.

References

[1] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):555–560, 2008.

[2] S. Ali and M. Shh. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[3] E. L. Andrade, B. Scott, and R. B. Fisher. Hidden markov models for optical flow analysis in crowds. In *Proceedings, of the International Conference on Pattern Recognition*, pages 460–463, 2006.

[4] A. Briassouli and I. Kompatsiaris. Robust temporal activity templates using higher order statistics. *IEEE Transactions on Image Processing*, 18(12):2756–2768, Dec. 2009.

[5] G. Brostow and R. Cipolla. Unsupervised bayesian detection of independent motion in crowds. In *Computer Vision and Pattern Recognition, 2006, IEEE Computer Society Conference on*, volume 1, pages 594–601, June 2006.

[6] A. B. Chan, M. Morrow, and N. Vasconcelos. Analysis of crowded scenes using holistic properties. In *IEEE Intl. Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2009)*, 2009.

[7] A. B. Chan and N. Vasconcelos. Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):909–926, 2008.

[8] H. Dee and D. Hogg. Detecting inexplicable behaviour. In *Proc. British Machine Vision Conference*, 2004.

[9] H. Dee and D. Hogg. Is it interesting? comparing human and machine judgements on the PETS dataset. In *Proc. Performance and Evaluation of Tracking Systems*, 2004.

[10] N. Doulamis. Evacuation planning through cognitive crowd tracking. In *16th International Conference on Systems, Signals and Image Processing, 2009. IWSSIP 2009*, pages 1–4, 2009.

[11] S. B. Ernesto, . Andrade, and R. B. Fisher. Modelling crowd scenes for event detection. In *Pattern Recognition, International Conference on*, pages 175–178, 2006.

[12] B. Ghanem and N. Ahuja. Phase based modelling of dynamic textures. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, 2007.

[13] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical Review*, 51(4282), 1995.

[14] A. B. W. . A. J. A. Jr. A look at motion in the frequency domain. In *NASA Technical Memorandum 84352*, 1983.

[15] L. Kratz and K. Nishino. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1446 – 1453, 2009.

[16] A. McIvor. Background subtraction techniques. In *Proceedings of the International Conference on Image and Vision Computing*, Auckland, New Zealand, 2000.

[17] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 101–108, 2008.

[18] E. S. Page. Continuous inspection scheme. *Biometrika*, 41:100–115, 1954.

[19] S. Pathan, A. Al-Hamadi, and B. Michaelis. Crowd behavior detection by statistical modeling of motion patterns. In *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of*, volume 1, pages 81–86, 2010.

[20] N. Vaswani, A. Chowdhury, and R. Chellappa. Activity recognition using the dynamics of the configuration of interacting objects. In *IEEE Conference Computer Vision and Pattern Recognition*, 2003.

[21] T. Xiang and S. Gong. Video behaviour profiling and abnormality detection without manual labelling. *International J. Computer Vision*, 67(1):21–51, 2006.

[22] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In *Proc. IEEE Conf. Computer Vision and Pattern Recongition*, 2004.