Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 8, August 2016, pg.254 – 261

Motion Tracking & Detection of Anomalous Events in the Crowd Based On Intelligence Video Surveillance

Liya R[1]

[1]Department of Computer Science and Engineering, Malabar CET, India

[1] liyakrshn@gmail.com

Abstract—Real-world actions occur often in crowded, dynamic environments. This poses a difficult challenge for current approaches to video event detection because it is difficult to segment the actor from the background due to distracting motion from other objects in the scene. The system proposes a technique for event recognition in crowded videos that reliably identifies actions in the presence of partial occlusion and background clutter. This approach is based on three key ideas: it efficiently match the volumetric representation of an event against over segmented spatio-temporal video volumes; it augment our shape-based features using flow; rather than treating an event template as an atomic entity, it separately match by parts (both in space and time), enabling robustness against occlusions and actor variability. The experiments based on human actions, such as picking up a dropped object or waving in a crowd show reliable detection with few false positives.

I. INTRODUCTION

With the In many applications, such as video surveillance, content based video coding, and human–computer interaction, moving object detection is an important and fundamental problem. The general technique for moving object detection is background elimination under the situation of fixed cameras. Detection of moving objects in video stream is the first related step of information removal in many computer visualization applications, including video surveillance, people tracking, traffic monitoring, and semantic annotation of videos. Video cameras are extensively used in surveillance application to examine public areas, such as train stations, airports and shopping centres. When crowds are intense, automatically tracking individuals becomes a difficult task. Anomaly detection is also known as outlier detection, which is applicable in a verity of application. Our proposed frame work is to implement a new tracking technique. Human monitoring is tiring, expensive, and ineffective. Our approach is a real time contribution to abnormal event detection and uses the motion of computational attenuation which quantifies motion saliency. The proposed method can be applied from small group of objects to dense touching moving objects like crowed. Crowded environments are very complicated to monitor by human observer, whether live or by means of video surveillance, because the optical patterns are highly recurring and the difficulty of the movement characterize the scene is often overpowering. Crowd finding is particularly significant in the background of intellectual and automated video surveillance systems proposed for large venues and public events, such as football games and concerts, as well as for such general environments as City Street and underground train stations during pinnacle hours. A crowd element can be defined as a region related to more than one person who has logical and identical motion. Crowd movement tracking is quite dissimilar from tracking individuals in the crowd. When individuals are being tracked, the

information is compute at the level of each individual. One purpose is to construct model of crowd performance and to detect irregular activities at the crowd level rather than at the individual level. Crowd analysis also finds applications in crowd simulation, crowd management, disaster management, outlet planning as well as other related areas. The purpose of this study is to analyse the crowd behaviour in real time in order to detect abnormalities that could lead to dangerous situations using computer vision and machine learning techniques.

II. BACKGROUND

A. Image processing

Image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging. Image processing allows one to enhance image features of interest while attenuating detail irrelevant to a given application, and then extract useful information about the scene from the enhanced image.

B. Global Descriptor

A video global descriptor is a set of features that describes the video as a whole and therefore is best able to describe the normal video patches. It is argued that classical handcrafted low-level features, such as HOG and HOF, may not be universally suitable and discriminative enough for every type of video. So that use low-level features, it use an unsupervised feature learning method based on auto-encoders. The auto-encoder learns sparse features based on gradient descent, by modelling a neural network.

C. Local Descriptors

To describe each video patch, it use a set of local features. The similarity between each patch and its neighboring patches are calculated. As for the neighbors, it consider nine spatial neighboring patches and one temporal neighboring patch (the one right behind the patch of interest when arranged temporally), yielding to 10 neighbors for each single patch. For temporal neighbors, it only consider the patch before the patch of interest (not the next one), even before the next video frames (and therefore patches) in the video stream arrive. It use SSIM for computing the similarity between two patches, which is a well-known image-quality assessment tool. Further, as a second type of local descriptor, to calculate the SSIM of each single frame with its subsequent frame in the patch of interest.

III. RELATED WORK

Extraction of human body in unconstrained still images [1]. Even though significant research has taken place on event and anomaly detection from static cameras [2], [3] the majority of these works address non-crowded scenes, where detailed visual information can be exploited for each individual. However, real-world surveillance scenarios often involve crowds of people or dense traffic, where such information cannot be easily extracted with traditionally used methods. Therefore, a number of different approaches have been proposed to handle these situations. Several interesting works [4]–[5] introduce tracking methods, nevertheless, they seem to be effective only in videos with crowds of low density, as tracking is otherwise hindered, due to the high degree of occlusions, while their computational cost is also greatly increased. As a result, the current SoA focuses mostly on analyzing entire frames spatially, temporally or both. Existing methods can be classified in two main categories: those that use only motion information to detect an abnormality in the scene, and those that use both appearance and motion information to describe the scene dynamics. In the first category, Wu et al. [6] use chaotic dynamics in particles' representative trajectories as a means to build a model capable of locating an outlier that moves with a different pattern. Even though this method works for very dense videos where a global motion pattern exists, it is unable to detect local abnormalities that take place in a small region in the frame, or in the absence of a global pattern. Activity recognition based exclusively on trajectories is also proposed by [7]. However, this method is only based on motion information, completely ignoring the existence of "interesting" activities that exhibit a typical motion pattern. In the same category, Mehran et al. [8] use the Social Force Model (SFM) to describe a crowd's normal behavior based on motion characteristics, while Cui et al. [9] make use of interaction energy potentials derived from the interest points' position and velocity. In [10], the min cut/max flow algorithm is used to define each block's dominant direction and crowd motion segmentation is

performed by training the algorithm separately, for each spatial location. That method also only relies on motion patterns to detect an anomaly, completely ignoring appearance information.

Another interesting work in the same domain is that of Cong et al. [11], who introduce a sparse reconstruction cost to measure the normality of the testing sample, considering dictionary learning methods. Saligrama and Chen [12] extract local low-level motion descriptors and utilize score functions for anomaly detection, derived from local nearest neighbor distances. A different approach is used by Adam et al. [13] who use fixed monitors to extract local low level features and determine a preset threshold for each monitor in order to declare an alert, while Kim and Grauman [14] propose a Markov Random Field model in a Bayesian framework for the final inference. A Gaussian mixture model is used by Ryan et al. [15] for anomaly detection in crowded scenes based on textures of optical flow in 3D volumes. 3D Gaussian distributions that characterize the underlying motion patterns of spatiotemporal cuboids are used in [16]. In this work, KL divergence is used as a distance measure to identify similar cuboids in the same location and new prototypes are created accordingly. Observations are then only evaluated according to distributions occurring in the same spatial location by creating a single HMM for each location. As a result the method proposed leads to many false positives in sparse videos, as the number of frames needed to work properly is huge, and it has to cover every region separately in order to train each HMM efficiently. Thus, that method is appropriate for crowded scenes of a very high density, but cannot handle videos of crowds with middle to low density, which are often captured by surveillance cameras. The same 3D gradient features are used by Lu et al. [17] in a different framework: they propose the use of sparse coefficients to fit new data to a previously learned dictionary. Sparse combination learning is introduced instead of searching the whole search space as classic sparsity- based methods do, thus greatly reducing the computational cost. The method exhibits remarkably high speed performance but at the expense of its accuracy, as shown in the experiments. Finally, an almost real-time algorithm is suggested in [18] for event detection, based on the clustering statistics derived from moving particles. A common problem that is encountered by all the methods mentioned earlier is their inability to successfully detect anomalies that move similarly to the "normal" motion pattern, as they rely solely on motion characteristics.

A second category of methods tackles this issue by incorporating appearance information as well. One work that stands out in this category is that of [19], that uses mixtures of dynamic textures to describe each 3D cuboids extracted from video sequence and detect temporal and spatial abnormalities. However, the computational cost of that algorithm, around 25 sec per frame, makes it prohibitive for many applications. An improved version of this method, with a lower computational cost, that is similar to ours, is found in [20]: that method's accuracy is also improved, but it still remains lower than ours as the experiments. The joint modeling of appearance and dynamics is also proposed by Ito et al. [21] for detecting interesting events via density estimation ratio to classify frames in two classes, normal or abnormal. Despite the applicability of that method to many scenarios, it is only suited for detecting events that occur over the entire frame (e.g. global changes in motions, scene changes etc.) and it misses local abnormalities. Another work that uses features based on both motion and texture is that of [22]. In that work, the input image is split into nonoverlapping cells and features based on motion, size and texture are extracted and are fed into two classifiers. The main drawback of the method is that the classification of each cell is determined by a pre-defined threshold, which makes the method sensitive to input video. Another interesting work is that of [23] which proposes a method of detecting abnormalities indirectly after establishing a complete interpretation of the foreground, by using a set of hypotheses. Afterwards, anomalies are defined as those hypotheses that are required to explain the foreground but which themselves cannot be explained by normal training samples. That method works efficiently on the UCSD dataset, however, the need for interpretation of all foreground objects may arise difficulties in more dense crowd datasets.

Results on the UCSD dataset provided also show that our method provides better accuracy. Finally, the work of [24] uses densely sampled spatiotemporal video volumes at each pixel location to construct a low level codebook and bag of video words is used to detect anomalous events. However, that method only uses the HOG descriptor to capture both motion and appearance characteristics omitting essential information that could led to better results. As the experiments show, our approach outperforms all the methods described above, in the important pixel level criterion on the UCSD dataset, which is used by most SoA works, making it more suitable for spatiotemporally local anomaly detection. The deployment of swarms involves the calculation of internal and external interaction forces, characterized by a number of parameters. Currently, new methods are being developed for the evaluation of parameter sensitivity in [25], which may be taken into account in future extensions of this work.

IV. PROBLEM DEFINITION

In this work, the problem of detecting dynamically changing anomalies in both space and time in videos with crowds of varying densities. In order to effectively capture these anomalies for a wide range of situations, the system incorporates both motion and appearance features. Our algorithm uses data derived from automatically extracted regions of interest (ROIs) instead of entire video frames, so as to only process pixels

containing information relevant to the event taking place, while at the same time achieving a lower computational cost, fewer false alarms, greater precision and successful spatiotemporal localization of anomalies, both on a global and local scale. In order to extract the ROIs, the system apply background subtraction using weighted moving mean, as it has been shown to be robust and reliable, however other SoA background subtraction methods like Gaussian Mixture Models (GMMs) could also be used, leading to equivalent results. The system define interest points on a dense grid in the resulting foreground and ROIs are described as rectangular areas of fixed size around each interest point. The size of the ROIs is determined at the beginning of each set of experiments, and depends on the camera viewpoint for each dataset. Due to the static nature of surveillance cameras, the block size needs to be set only once for each camera, or in our case for each dataset, and thus does not affect our algorithm's generality. For the UCSD dataset, a ROI of 20×20 pixels is used, as it is large enough to capture activity/appearance related details, but is not too large, so as to include noisy information in the descriptor. Once ROIs are extracted, the interest points in them are tracked until the next frames using the KLT tracker, while the foreground grid is continuously updated, with new interest points defined in each new frame's foreground area. The resulting ROIs and the interest points in them are considered informative and are retained if at least 60% of that ROI contains motion, otherwise that interest point and its ROI are considered to be noisy and are ignored. The ROI needs to contain at least 60% moving pixels in order to be as informative as possible; if a ROI contains fewer moving pixels, noisy (motionless) data will also be taken into account, while if it is required to contain more moving pixels, potentially informative interest points may be ignored. Spatiotemporal feature extraction from ROIs follows for a particular time window, to acquire descriptors that effectively describe the video dynamics, and help identify both local and global abnormalities. The system considers both motion and appearance features, as their combined use allows the detection of anomalies, i.e. deviations of motion and/or appearance from usual patterns, leading to a generally applicable method. An overview of the procedure for extracting the descriptor is depicted in Fig.1. The stages for modelling appearance and motion are discussed in more detail in the sequel.

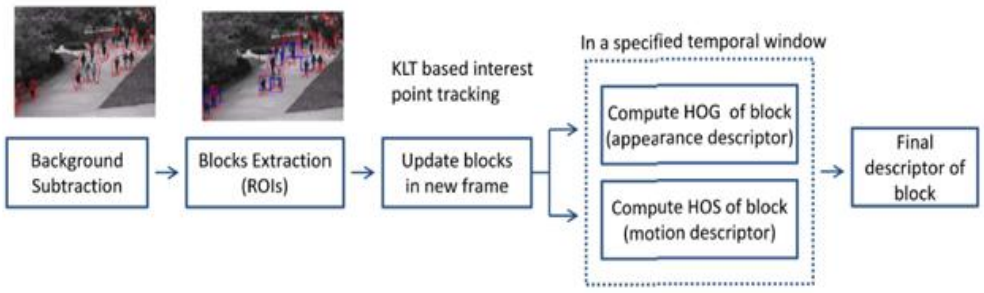

Fig.1. Overview of final motion-appearance descriptor calculation.

A. Appearance Modelling

The vector space model allows representation of document as a vector. If a term present in the file, its value in the vector is nonzero, otherwise is zero. A query is also represented as a vector and if the term is queried, the dimension assigned as 1, 0 if not queried. The (tf-idf) weighting involves two components: Term frequency and inverse document frequency. Term frequency denotes the number of occurrences of term t in file f. The inverse document frequency is idf = log (N/df) where N denotes the total number of files and df is the number of files that contains the term t. In order to extract the appearance characteristics of a video sequence, the Histograms of Oriented Gradients (HOG) proposed in [26] are used, as the HOG descriptor has several advantages over other appearance features: it is color invariant as it uses gray scale images, and is also invariant to illumination and local geometric transformations as a result of the normalization that takes place. At the same time, it effectively captures the local edge and gradient structure, so it can distinguish variations in appearance even in small areas of the image. The implementation of HOG that is adopted is that of [27], as it creates direction invariant HOGs by following a mirroring technique, where mirrored shapes are mapped into the same bin. Direction invariant appearance features (HOGs) decrease intra-class variation, e.g. for walking, which is the predominant activity in human crowds, resulting in similar appearance descriptors for motions in opposite directions. This leads to more robust appearance descriptors that are suitable for the needs of anomaly detection in crowded videos, which can describe, for example, the density or sparseness of a crowd more effectively by ignoring directionality (which is not relevant for appearance). The HOG descriptor is applied in ROI blocks that are tracked over time and are extracted as described in the previous section, so the final HOG descriptor for each block also incorporates temporal information. The procedure for this computation is as follows: each block k is first divided into 2×2 cells as suggested in [26] for a more detailed description that also takes spatial location information into account and mitigates the effects of local noise. For example, if occlusions are present in a ROI,

its division into 2×2 cells may limit their presence to only one of the cells, instead of the whole area, leading to a less noisy appearance descriptor. The division of a block into 2×2 cells was chosen, as it was found by Dalal and Triggs [26] to retain a sufficient level of detail for describing appearance. A weighted histogram of gradients is then created for each cell using 9 bins, corresponding to the gradients' orientation. The HOG of the cth cell ($1 \leq c \leq 4$) in block k of frame j is thus represented by $HOG_j^k(c)$, of dimensions 1×9. Each histogram is normalized and the 4 resulting cell histograms are concatenated, forming a 1×36 block descriptor, which is also normalized for noise elimination. Once HOGs for each block are calculated for all frames in the temporal window under examination, they are averaged over 3 consecutive frames so as to include richer temporal information and at the same time achieve temporally local noise reduction. The final appearance descriptor is thus a concatenation of a 3 frame average for each cell c in block k:

$$\overline{HOG}_{j,j+2}^k(c) = E\left[HOG_j^k(c), HOG_{j+1}^k(c), HOG_{j+2}^k(c)\right] \quad (1).$$

This means that a 15 frame time window will result in 5 concatenated triplets of 1×36 descriptors, resulting in a 1×180 final spatiotemporal appearance descriptor. The entire process for extracting HOG descriptor is depicted in Fig.2.

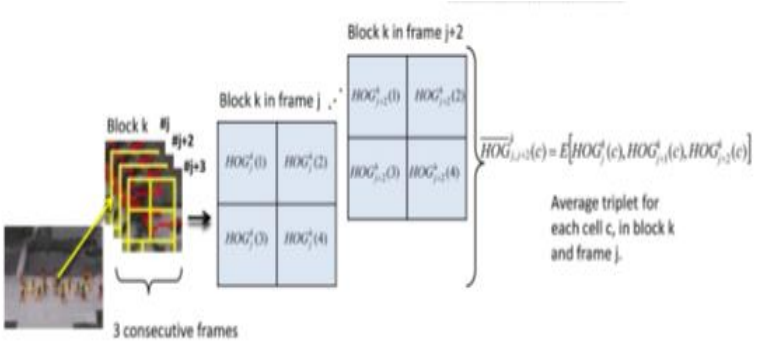

Fig.2. Extraction of appearance descriptor (HOG)

For simplicity of notation, in the sequel, the HOG descriptor of Eq. (1) for block k, averaged over frames j to j +2 will be represented as $\overline{HOG}_{j,j+2}$ including the average over all 4 cells.

B. Motion Modelling Using HOS Descriptor

This work introduces a novel method for capturing crowd dynamics based on the application of swarm intelligence, which is used to build a novel motion descriptor. Swarm intelligence in computer science is inspired from the behaviour and characteristics of real swarms encountered in nature. Swarms are comprised of individuals, which act autonomously, while following the specific rules of a swarm and interacting with each other. Although the decisions of a swarm's individuals take place locally, their aggregated behaviour can match events in crowded environments, which makes them relevant in many applications. Swarm based methods have been used in the literature for image filtering and noise reduction [28], but their incorporation for the analysis of motion in videos is an original concept first presented in [29]. The core idea is the monitoring of movements in crowded scenes by a swarm of agents "flying" over them, to capture their dynamics in a collective way while also taking motion history into account. Swarms are thus deployed and the agents' positions are extracted from their accelerated motion, derived from the forces acting on the swarm. They are then used to form Histograms of Oriented Swarms (HOS), which are used to capture the ROIs' underlying motion and detect anomalous events in them. The main concepts of our swarm descriptor are presented in the following section.

V. SWARM MODELLING FOR CROWDS DYNAMICS

The overview of the proposed secure search system model is given in figure 1. In our implementation, the system adopts physics-based modelling of crowded scenes, as their properties are highly correlated with those of a swarm in nature. The swarm model that is used is based on the general theory described in [28] and on the behavior of natural swarms, consisting of predators, which "fly" over the "prey", following its dynamics. In [28], swarm modelling is used to filter noise in images, whereas in this work it is deployed to better characterize the highly complex and stochastic motion information from videos of crowds. In our implementation, swarms comprise of agents and a prey: the agents "track" the prey, but also interact with each other, as they would in nature. Hence, agents ("predators") are subject to three types of forces: "physical" forces, like inertia and friction, interaction forces between them, and external forces dependent on the prey. Interaction forces ensure the cohesion of the swarm of agents, friction forces maintain elementary memory of the agents' velocity, while external forces depend on the characteristics of the prey being tracked. Consequently, in this approach, swarm intelligence maps the motion information into a more informative space by efficiently tracking

the motion represented by the prey. Agents filter the prey motion, avoiding false alarms and local noise caused e.g. by occlusions or outlier optical flow values. The prey corresponds to the values of the variable that system wants to leverage in the discriminative process. In our case, the extraction of motion features via the swarm modelling, so optical flow (OF) values are used as a prey, as detailed in the next section. Thus, the use of swarms is expected to lead to better results than when using OF information alone, as they can capture the most important aspects of crowd behaviour while circumventing the effects of local noise, occlusions and the overall complexity of motion in crowded scenes.

A. Prey Generation

The prey that is tracked by the swarm comprises of magnitude values of pixels lying inside ROIs, instead of their luminance, which is the case in [29]. Hence, the number of prey in each frame varies, as it is equal to the number of ROIs in the frame. In this section describe how prey data is extracted, namely how it is mapped to be tracked by agents.

B. Extraction of forces

The manner in which the agents operate, i.e. the way they "fly over" the prey and track it. Agents are groups that the system defines to track the prey and characterize its state: they are initially located in random positions, which change over time according to agent-prey forces, agent-to-agent forces and friction forces presented here. The result of these forces' interactions is the accelerated motion of the agents, which is affected and formed according to prey behaviour. These forces are inspired by crowd psychology and the analysis of movements of individuals in crowds, matching real world behaviors of people (or other entities, like cars or animals) in crowded situations: for example, when agents are too close to each other, repulsive forces develop between them, while the opposite occurs (attraction forces develop) when they are at a large distance, ensuring the cohesion of the swarm of agents.

VI. ANOMALY DETECTION AND LOCALIZATION

Appearance and motion descriptors are combined to form the final descriptor for anomaly detection. In a time window of m frames, average triplets of HOG and HOS are consecutively concatenated. The overall process takes place in each ROI and it is depicted in Fig. 3. A normalization step takes place to form the final descriptor so as to achieve scale invariance.

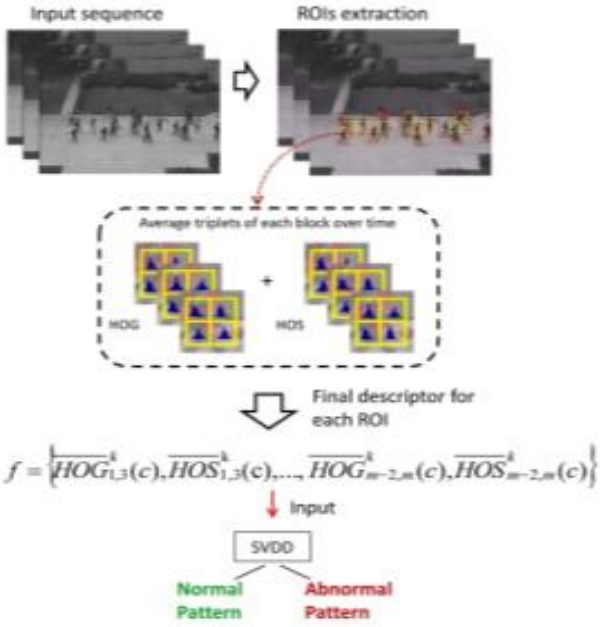

Fig.3. Overview of method proposed in a time window of m frames.

A Support Vector Machine (SVM) is used to determine each region's normality. SVMs are used, as they generally exhibit good performance relatively to other machine learning methods and are also fast to run, for reliable real time detection. Furthermore, they are able to handle large data sets, which generally appear in real life situations. Because of the infinite number of "anomalies" that can be derived in each case, it is impossible to provide examples of all possible anomaly classes, so a one class classifier is chosen. This way, we provide our system exclusively with normal situations, aiming to identify any irregularities deviating from the normal pattern. This leads to a more accurate and general classifier capable of detecting different kinds of anomalies, even when appearing for the first time in the dataset.

VII. FUTURE WORKS

In future algorithms can be implemented for most challenging pixel level criterion, challenging crowd videos with many occlusions, local noise and local scale variations. This fact in combination with its low computational cost and its effectiveness in different environments, make our algorithm further very appropriate for a variety of surveillance applications.

VIII. CONCLUSIONS

In this work, the system proposes a novel framework for anomaly detection in different scenarios, recorded from static surveillance cameras. Swarm intelligence is exploited for the extraction of robust motion characteristics and together, with appearance features, form a descriptor capable of effectively describing each scene. Its remarkable performance in 4 completely different kinds of datasets proves the method's generality and its applicability in real life situations. The high detection rate in the UCSD dataset, that greatly outperforms various state-of-the-art approaches.

REFERENCES

[1] Athanasios Tsitsoulis, Member, IEEE, and Nikolaos G. Bourbakis, Fellow, IEEE, "A Methodology for Extracting Standing Human Bodies From Single Images", vol.45, no.3, june 2015.

[2] Athanasios Tsitsoulis, Nikolaos G. Bourbakis, "A methodology for extracting standing human bodies from single images," IEEE Trans. Image Process., vol. 45, no. 3, pp. 2168-2291, June 2015

[3] N. P. Cuntoor, B. Yegnanarayana, and R. Chellappa, "Activity modeling using event probability sequences," IEEE Trans. Image Process., vol. 17, no. 4, pp. 594–607, Apr. 2008.

[4] L. Wang, Y. Qiao, and X. Tang, "Latent hierarchical model of tem- poral structure for complex activity classification," IEEE Trans. Image Process., vol. 23, no. 2, pp. 810–822, Feb. 2014.

[5] W. Hu, X. Zhou, W. Li, W. Luo, X. Zhang, and S. Maybank, "Active contour-based visual tracking by integrating colors, shapes, and motions," IEEE Trans. Image Process., vol. 22, no. 5, pp. 1778–1792, May 2013.

[6] D. Tran, J. Yuan, and D. Forsyth, "Video event detection: From sub- volume localization to spatiotemporal path search," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 2, pp. 404–416, Feb. 2014.

[7] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2010, pp. 2054–2060.

[8] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques, "Activity recognition using a mixture of vector fields," IEEE Trans. Image Process., vol. 22, no. 5, pp. 1712–1725, May 2013.

[9] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 935–942.

[10] X. Cui, Q. Liu, M. Gao, and D. N. Metaxas, "Abnormal detection using interaction energy potentials," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2011, pp. 3161–3167.

[11] H. Ullah and N. Conci, "Crowd motion segmentation and anomaly detection via multi-label optimization," in Proc. IEEE Int. Conf. Pattern Recognit. Workshop (ICPRW), Nov. 2012.

[12] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnor- mal event detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2011, pp. 3449–3456.

[13] V. Saligrama and Z. Chen, "Video anomaly detection based on local statistical aggregates," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2012, pp. 2112–2119.

[14] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real- time unusual event detection using multiple fixed-location monitors," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 3, pp. 555–560, Mar. 2008.

[15] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 2921–2928.

[16] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "Textures of optical flow for real-time anomaly detection in crowds," in Proc. 8th IEEE Int. Conf. Adv. Video Signal-Based Surveill. (AVSS), Aug./Sep. 2011, pp. 230–235.

[17] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2009, pp. 1446–1453.

[18] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2013, pp. 2720–2727.

[19] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis, "Timely, robust crowd event characterization," in Proc. 19th IEEE Int. Conf. Image Process. (ICIP), Sep./Oct. 2012, pp. 2697–2700.
[20] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2010, pp. 1975–1981.
[21] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 1, pp. 18–32, Jan. 2014.
[22] Y. Ito, K. Kitani, J. Bagnell, and M. Hebert, "Detecting interesting events using unsupervised density ratio estimation," in Proc. IEEE Conf. Eur. Conf. Comput. Vis. Workshop (ECCVW), vol. 7585. Oct. 2012, pp. 151–161.
[23] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2011, pp. 55–61.
[24] B. Antic and B. Ommer, "Video parsing for abnormality detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Nov. 2011, pp. 2415–2422.
[25] M. J. Roshtkhari and M. D. Levine, "Online dominant and anomalous behavior detection in videos," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2013, pp. 2611–2618.
[26] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1. Jun. 2005, pp. 886–893.
[27] K. Avgerinakis, A. Briassouli, and I. Kompatsiaris, "Recognition of activities of daily living for smart home environments," in Proc. 9th Int. Conf. Intell. Environ. (IE), Jul. 2013, pp. 173–180.
[28] H. M. H. Teodorescu and D. J. Malan, "Swarm filtering procedure and application to MRI mammography," Polibits, vol. 42, no. 42, pp. 59–64, 2010.
[29] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis, "Swarm- based motion features for anomaly detection in crowds," in Proc. IEEE Int. Conf. Image Process. (ICIP), Oct. 2014, pp. 2353–2357.

