Chapter 15
Anomaly Detection in Crowded Scenes: A Novel Framework Based on Swarm Optimization and Social Force Modeling

R. Raghavendra, M. Cristani, A. Del Bue, E. Sangineto, and V. Murino

Abstract This chapter presents a novel scheme for analyzing the crowd behavior from visual crowded scenes. The proposed method starts from the assumption that the interaction force, as estimated by the Social Force Model (SFM), is a significant feature to analyze crowd behavior. We step forward this hypothesis by optimizing this force using Particle Swarm Optimization (PSO) to perform the advection of a particle population spread randomly over the image frames. The population of particles is drifted towards the areas of the main image motion, driven by the PSO fitness function aimed at minimizing the interaction force, so as to model the most diffused, normal behavior of the crowd. We then use this proposed particle advection scheme to detect both global and local anomaly events in the crowded scene. A large set of experiments are carried out on public available datasets and results show the consistent higher performances of the proposed method as compared to other state-of-the-art algorithms.

15.1 Introduction

Recently, major research efforts are underway in the computer vision community to develop robust algorithms for understanding the behavior of crowds in video surveillance contexts. Anomaly detection in crowded scenes is an important social problem far from being reliably solved. This is because conventional methods designed for surveillance applications fail drastically for the following reasons: (1) severe overlapping between individual subjects; (2) random variations in the density of people over time; (3) low resolution videos with temporal variations of

R. Raghavendra • M. Cristani • A. Del Bue • E. Sangineto • V. Murino (✉)
Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia,
via Morego 30, 16163 Genova, Italy
e-mail: vittorio.murino@iit.it

the scene background. Nowadays, crowds are viewed as the very outliers of the social sciences [27]. Such an attitude is reflected by the remarkable paucity of psychological research on crowd processes [27].

The main objective of crowd behavior analysis involves not only modeling of people mass dynamics but also detecting or even predicting possible abnormal or anomalous behaviors in the scene. In particular for surveillance scenarios, this task is of paramount importance since early detection, or even prediction, may reduce the possible dangerous consequences of a threatening event, or may alert a human operator for inspecting more carefully the ongoing situation.

Anomaly detection in crowded scenes can be classified into two types: (1) local abnormal event, indicating that a behavior in a specific local image (or frame) area is different from that of its neighbors in spatio-temporal terms; (2) global abnormal event, indicating that the whole frame is abnormal irrespective of the local regions. In other words, a global abnormal event detection aims at classifying each frame as either abnormal or normal, while in local detection we also want to localize the parts of the given frame which likely contain the abnormal activity.

In this article we present both global [26] and local [25] anomaly detection techniques which have been tested on different real-time scenarios. We developed these techniques based on the assumption that people in the crowd behave in ways like birds (also known as particles) in a swarm. Thus, we try to address crowd behavior analysis by considering the crowd as mutually interacting birds in a swarm.

In general, a crowd can be considered as a collection of mutually interacting people, where random individuals' motion, due to the influence of neighbors, spatial physical structure of the scene, etc., will dominate the dynamics and the flow of the crowd. With this primary idea, we make an attempt to reflect a visual crowd behavior using the concept of Swarm Optimization. Typically, the idea of Swarm Optimization derived from the flight control (defined by a fitness function) of randomly dispersed birds (also referred to as particles) in a given space. In this framework, both local and social behavior among the birds or particles in the swarm is considered. Similarly, we represent people in a crowd as interacting particles following an evolutionary dynamic. These dynamics are driven by a fitness function and they are influenced by the interaction forces among the swarm particles. With this motivation, we propose a novel framework for particle advection using PSO [15] and Social Force Model (SFM) [13]. The proposed method belongs to the class of particle advection schemes and it is based on the assumption that the evolving interaction forces estimated using SFM is a significant feature for analyzing the crowd behavior. Our scheme starts by initializing particles randomly on the initial video frame, which are then optimized and drifted to the main regions of the motion according to a fitness function suitably defined. The aim of the fitness function is to minimize the interaction forces, so as to model the most diffused, normal behavior of the crowd as suggested by behavioral studies. Hence, the anomalies are identified by the particles whose force significantly deviates from the typical force magnitude.

We put forward this framework to detect two different kinds of anomalies namely: global and local anomalies. In order to detect global anomalies, we process the interaction force obtained using the PSO-SFM method by detecting the change in its magnitude. On the other hand, local anomaly detection is carried out by checking if some particles (i.e., their interaction forces) do not fit the estimated "typical" distribution, and this is done using a RANSAC-like method followed by a segmentation algorithm to finely localize the abnormal areas.

There are several characteristics which differentiate our approach with respect to other related works. First, particles are spread randomly over the image and can move in a continuous way according to an optimization criterion, differently from other approaches which constrain the particles in a priori fixed grid. Second, we use PSO for particle advection which considers not only the individual particles motion, but also the global motion of the particles as a whole, i.e., social interactions.

Extensive experiments are carried out on different types of public available video datasets to prove the effectiveness of the proposed scheme. In order to evaluate the global anomaly scheme, we considered four different public available datasets, namely: UMN, PETS 2009, UCF and also a challenging dataset that reflects the prison riots, download by YouTube. In order to evaluate the proposed scheme for local anomaly detection, we consider two different public datasets, namely UCSD and MALL datasets.

The rest of this chapter is organized as follows: Sect. 15.2 shows the state-of-the-art techniques for crowd behavior analysis from the computer vision point of view. Section 15.3 describes the proposed particle advection approach based on the PSO-SFM model and also discusses the global and local anomaly detection schemes. Section 15.4 presents the experimental results. Finally, Sect. 15.5 draws the conclusions.

15.2 Related Work

Several techniques have been proposed for the anomaly detection in visually crowded scenes. State of the art methods can be coarsely classified into two different types: model-based and particle advection-based approaches. Among these two methods, the particle advection based approaches will more naturally represent the holistic view of a crowd and they do not require the segmentation or detection of individuals. On the contrary, the outcome of these algorithms may eventually result in the detection of individuals when they are detected as an anomaly. Here, we first review the literature on model based approaches which is then followed by particle advection schemes.

15.2.1   Model Based Approaches

In [29], a novel unsupervised framework is presented to model the pedestrian activities and interactions in crowded scenes. Here, low level visual features are computed by carrying out the intensity difference between successive frames of a given video. Then, these low level features are labeled using their location and motion direction to form a basic feature set. The features are then quantized into visual words to construct a dictionary. Finally, the activities are classified using two well know classifiers namely: Latent Dirichlet Allocation (LDA) mixture model and Hierarchical Dirichlet Process (HDP) mixture model.

In [20], a dynamic texture model is employed to jointly model the appearance and dynamics of the crowded scene. This method explicitly addresses the detection of both temporal and spatial anomalies. Further, a new dataset of crowded scenes with videos of the walkway of a college campus and crowd with naturally varying densities are made available for the vision community. In [17], steady state motion of the crowd behavior is exploited by analyzing the underlying structure formed by the spatial and temporal variations in the motion. Then, a Hidden Markov Model (HMM) is trained on the motion patterns at each spatial location of the video to predict the motion pattern that is exhibited by the subjects as they transverse through the video. Finally, anomalous activities are detected as low likelihood motion patterns.

In [16], anomaly detection in the crowded scene is carried out using a space-time Markov Random Field (MRF) model. Given a video, a MRF graph is constructed by dividing each frame into a grid of spatio-temporal local regions. Each region corresponds to a single node and neighboring nodes are connected with links. Then, each node is associated with an optical flow observation to learn the atomic motion patterns using a mixture of probabilistic principal component analysis. Finally, inference on the graph is carried out to decide whether each node is normal or abnormal. In [1], a histogram is used to measure the optical flow probability in local patterns of the image and then an ambiguity based threshold is selected to monitor and detect the anomalies in the input videos. Further, a new video dataset with different anomaly scenarios is made available to the vision community. In [3], a new technique based on video parsing is proposed for accurate abnormality detection in the visual crowded scene. Each video frame is parsed by establishing a set of hypotheses that jointly provide information on the entire foreground. Finally, a probabilistic model is employed to localize the abnormality using statistical inference. In [18], dense optical flow fields are computed between two successive frames to obtain the low level motion information in terms of direction and magnitude for each pixel. Then, 2D histograms of motion direction and magnitude for all flow vectors are computed. A symmetry measure is computed by summing the absolute difference between the 2D histogram and a flipped version of itself to determine the anomaly in the scene. Extensive experiments are carried out on the LoveParade 2010 dataset to prove the reliability of the method. In [9], a sparse reconstruction cost is proposed to detect the presence of anomalies in

crowded scenes. Here local spatio-temporal patches are used to construct the normal dictionary. Further, to reduce the size of the dictionary, a new selection method is proposed based on sparsity consistency constraints.

15.2.2   Particle Advection Based Approaches

In case of particle advection schemes, a grid of particles is usually considered in each frame which are then advected using the underlying motion data [2,21,22,30]. The assumption here is that each particle is considered as an atomic entity in the mass of people, and the trajectories generated from the particles' advection may portray significant information concerning representative properties of the scene in terms of both characteristics of the physical area and the crowd behavior. The first work using particle advection schemes for crowd behavior analysis was introduced in [2]. Here, the particle flow is computed by moving a grid of particles using the fourth-order Runge-Kutta-Fehlberg algorithm [19] along with the bilinear interpolation of the optical flow field. This method is further extended in [30] using chaotic invariants capable of analyzing both coherent and incoherent scenes. In [22], streaklines are introduced and integrated with a particle advection scheme capable of incorporating the spatial change in the particle flow.

In [21] the social force model (SFM) [13] is exploited to detect abnormal events. After the superposition of a fixed grid of particles on each frame, the SFM is used to estimate the interaction force. In turn, the interaction force is used to describe (abnormal) crowd behavior. So, after estimating the so-called force flow, a bag of words method [4] and a Latent Dirichlet Allocation (LDA) [5] are employed to discriminate between normal and abnormal frames. Possible abnormal areas are localized selecting those regions with the highest force magnitude. In [23] the authors provide an excellent analysis of the above mentioned particle advection schemes in which crowd is dealt with using hydrodynamics principles.

15.2.3   Discussion

In Fig. 15.1a we show the result obtained applying the state-of-the-art people detector of Dalal and Triggs [11] to a crowd image. Only 5 out of 23 persons are correctly detected. Moreover, two false positives (the big rectangles) are also included in the outcome. The situation is even worse in the densely crowded image shown in Fig. 15.1b, where the automatic people detection phase clearly fails in localizing the huge number of persons here represented. These two examples show why approaches based on detection or segmentation of individuals are barely robust when applied to the analysis of non-sparsely crowded scenes.

Conversely, particle advection methods do not rely on people segmentation and assume that a crowd can be represented by a set of particles influenced by the

Fig. 15.1 Examples of common people detector errors on a low-crowded (**a**) and a high-crowded (**b**) scenario. The large number of false positives and false negatives makes the use of people detector-based techniques highly unreliably for crowd analysis

people's movements. The particles' flow is then analyzed trying to detect possible anomalies. In Sect. 15.3.5 we will show that our anomaly detection approach is able to localize an anomaly in the frame shown in Fig. 15.1a (i.e., a man on a bicycle with a velocity higher than the surrounding pedestrians). In fact, we can detect the person(s) in the scene with an anomaly behavior by back-projecting the particle positions corresponding to the localized anomaly into the image.

Before concluding this section, we refer the reader interested in crowd behavior analysis details to recent review papers. In [31], a survey on available techniques for crowd modeling from both the computer vision and the crowd simulation point of view are presented. Emphasis is drawn on discussing the techniques available for crowd modeling using agent based models, nature based models and physical models. In [14] a discussion on the available computer vision techniques for crowd behavior analysis for video surveillance applications is presented. This survey also reports a few computer vision schemes able to address problems like crowd dynamics, crowd analysis and crowd synthesis. In [10] a summary of crowd behavior techniques from a social signal perspective applied to video surveillance is presented.

15.3 Proposed Particle Advection Using PSO-SFM

This section describes our proposed particle advection method using PSO-SFM. In earlier attempts [2, 21], the particle advection is carried out by placing a rectangular grid of particles over each video frames. Then, the velocity for each particle is calculated using fourth-order Runge-Kutta-Fehlberg algorithm [19] along with the bilinear interpolation of the optical flow field. In general, a drawback of this

approach is that it assumes that a crowd follows a fluid-dynamical model which is too restrictive when modeling masses of people. The elements of the crowd may also move with unpredictable trajectories that will result in an unstructured flow. Moreover, the use of a rectangular grid for particles is a coarse approximation with respect to the continuous evolution of the social force. To overcome these drawbacks, we propose a novel particle advection scheme using PSO aiming at modeling the crowd behavior. Before presenting the detailed description of our proposed scheme, we first provide a brief introduction on PSO and SFM in the following subsections.

15.3.1 *Particle Swarm Optimization*

Particle Swarm Optimization is a stochastic, iterative, population-based optimization technique aimed at finding a solution to an optimization problem in a search space [15]. The main objective of PSO is to optimize a given criterion function called fitness function f. PSO is initialized with a population, namely a *swarm*, of N-dimensional particles distributed randomly over the search space (of dimension N too): each particle is so considered as a point in this N-dimensional space and the optimization process manages to move the particles according to the evaluation of the fitness function in an iterative way. More specifically, at each iteration, each particle is updated according to two "best" values, respectively called $pbest_i$, which depends on the i-th particle, and $gbest$ which is independent from the specific particle. $pbest_i$ is the position corresponding to the best (e.g., minimum) fitness value of particle i obtained so far (i.e. taking into account the positions computed from the first iteration to the current one). On the other hand, $gbest$ is the best position achieved by the whole swarm:

$$gbest = \arg\min_i f(pbest_i), \tag{15.1}$$

The position change (called "velocity") v_i for the i-th particle is updated according to the following equations [15]:

$$v_i^{new} = I_A \cdot v_i^{old} + C_1 \cdot rand_1 \cdot (pbest_i - x_i^{old})$$
$$+ C_2 \cdot rand_2 \cdot (gbest - x_i^{old}); \tag{15.2}$$
$$x_i^{new} = x_i^{old} + v_i^{new}, \tag{15.3}$$

where I_A is the inertia weight, whose value should be tuned to provide a good balance between global and local explorations, and it may result in fewer iterations on average for finding near optimal results. The scalar values C_1 and C_2 are acceleration parameters used to drive each particle towards $pbest_i$ and $gbest$. Low values of C_1 and C_2 allow the particles to roam far from target regions, while high

values result in abrupt movements towards the target regions. $rand_1$ and $rand_2$ are random numbers between 0 and 1. Finally, x_i^{old} and x_i^{new} are the current and updated particle positions, respectively, and the same applies for the deviation v_i^{old} and v_i^{new}.

15.3.2 Social Force Model

The SFM [13] provides a mathematical formalization to describe the movement of each individual in a crowd on the basis of its interaction with the environment and other obstacles. The SFM can be written as:

$$m_i \frac{dW_i}{dt} = m_i \left(\frac{W_i^p - W_i}{\tau_i} \right) + F_{int}, \qquad (15.4)$$

where m_i denotes the mass of the individual, W_i indicates its *actual velocity* which varies given the presence of obstacles in the scene and τ_i is a relaxing parameter. F_{int} indicates the interaction force experienced by the individual which is defined as the sum of attraction and repulsive forces. Finally, W_i^p is the *desired velocity* of the individual.

Assuming $m_i = 1$ and $\tau_i = 1$, from Eq. (15.4) we obtain:

$$F_{int} = W_i - W_i^p + \frac{dW_i}{dt}. \qquad (15.5)$$

Equation (15.5) shows that the higher the difference between the actual and the desired velocities of a particle, the stronger its interaction force. The intuitive idea behind this is that an obstacle (e.g., a person or a group of persons) can make a particle (representing an individual of the analyzed crowd) to deviate from its desired path. The higher this deviation, the stronger the underlying interaction force. Thus, estimating the interaction force of the particle swarm will give us an instrument to assess the total amount of person-to-person interactions in a given frame. Anomalies will be detected as outliers in the interaction force distribution.

In the next section we will see how the optical flow can be used for an operational definition of the velocities involved in Eq. (15.5) and the how the PSO process can be used to simulate the movement of a set of individuals who aim at minimizing their respective interaction forces.

15.3.3 The Proposed Minimization Scheme

The PSO begins with a random initialization of the particles in the first frame. From such initial stage, we obtain a first guess of *pbest_i*, for each particle i, and the

global *gbest*. The particles are defined by their 2-D positions corresponding to the pixel coordinates in the frames. At each iteration, the *pbest$_i$* value is updated only if the present position of the particle is better than the previous position according to fitness function evaluated on the model interaction force. Finally, the *gbest* is updated with the position obtained from the best *pbest$_i$* after reaching the maximum number of iterations or if the desired fitness value is achieved. We then use the final particle positions as the initial guess in the next frame and the same iterative process is repeated until the end of the video sequence. Therefore, the movement of the particles is updated according to the fitness function which drives the particles toward the areas of minimum interaction force using SFM.

15.3.3.1 Computing the Fitness Function

The fitness function aims at capturing the interaction force exhibited by each movement in the crowded scene. Each particle is evaluated according to its interaction force calculated using SFM and optical flow [6]. In fact, the Optical Flow (OF) is a good candidate to substitute the pedestrian velocities in the SFM model.

Using OF, we define the actual velocity of particle *i* as:

$$W_i = O_{avg}(x_i^{new}), \tag{15.6}$$

where $O_{avg}(x_i^{new})$ indicates the average OF at the particle coordinates x_i^{new}, which in turn is estimated using Eq. (15.2). The average is computed over L previous frames. The desired velocity of the particle is defined as:

$$W_i^p = O(x_i^{new}), \tag{15.7}$$

where $O(x_i^{new})$ represents the OF intensity (in the current frame) of the particle *i*. Both $O()$ and $O_{avg}()$ are computed using interpolation in a small spatial neighborhood to avoid numerical instabilities of the OF. Finally, we calculate the interaction force F_{int} using Eq. (15.5):

$$F_{int}(x_i^{new}) = \frac{dW_i}{dt} - \left(W_i^p - W_i\right), \tag{15.8}$$

where the velocity derivative is approximated as the difference of the OF at the current frame t and $t - 1$, that is $\frac{dW_i}{dt} = [O(x_i^{new})|_t - O(x_i^{new})|_{t-1}]$. As above mentioned, the interaction force (Eq. (15.5)) allows an individual to change its movement from the desired path to the actual one. This process is in some way mimicked by the particles which are driven by the OF toward the image areas of larger motion. In this way, the more regular the pedestrians' motion, the less the interaction force, since the people motion flow varies smoothly. So, in a normal crowded scenario the interaction force is expected to stabilize at a certain (low) value
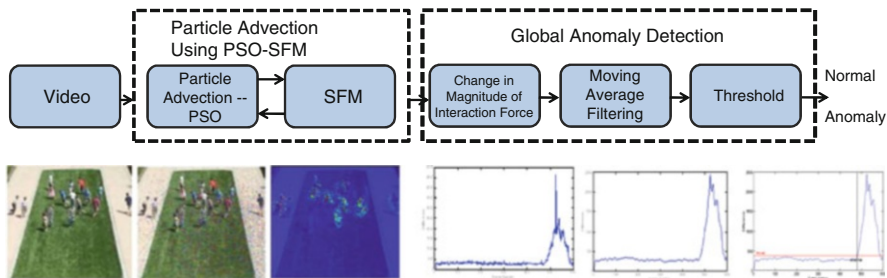
Fig. 15.2 Block diagram of the proposed framework for global anomaly detection

complying with the typical motion flow of the mass of people. It is then reasonable to define a fitness function aimed at minimizing the interaction force and moving particles toward these sinks of small interaction force, thereby allowing particles to simulate a "normal" situation of the crowd.

Hence, we define our fitness function as:

$$f(x_i) = F_{int}(x_i), \qquad (15.9)$$

where x_i denotes the i-th particle's position. With the above definitions we can use the PSO framework presented in Sect. 15.3.1 to minimize $f()$.

15.3.4 Global Anomaly Detection Scheme Using PSO-SFM

In Fig. 15.2 we show the stages of our global anomaly detection system, whose aim is to classify every frame of a given video sequence as either "normal" or "abnormal". In the first stage we estimate the interaction force on each frame using the PSO-SFM scheme described in Sect. 15.3.3. The interaction force associated with each particle is then processed further to identify the global anomaly in the frame.

As an example, Fig. 15.3a–d show the computed interaction force with the proposed particle advection using PSO-SFM for both normal (Fig. 15.3a, b) and anomaly video frames (Fig. 15.3c, d). In these figures, we plotted on the image the magnitude of the interaction forces assigned to every particle. As observed in Fig. 15.3, the presence of the high magnitude interaction force over time can provide useful information about the existence of an anomaly. This allow us to formulate the detection of global anomalies as the detection of the changes in the interaction force magnitude. This process is valid with the proposed particle advection scheme since the presence of global abnormality can be recognized by the presence of high magnitude of the interaction force associated with the particles (see Fig. 15.3). Since all the available test videos contains a certain amount of frames in which normal behavior is assumed, we take advantage of this information in
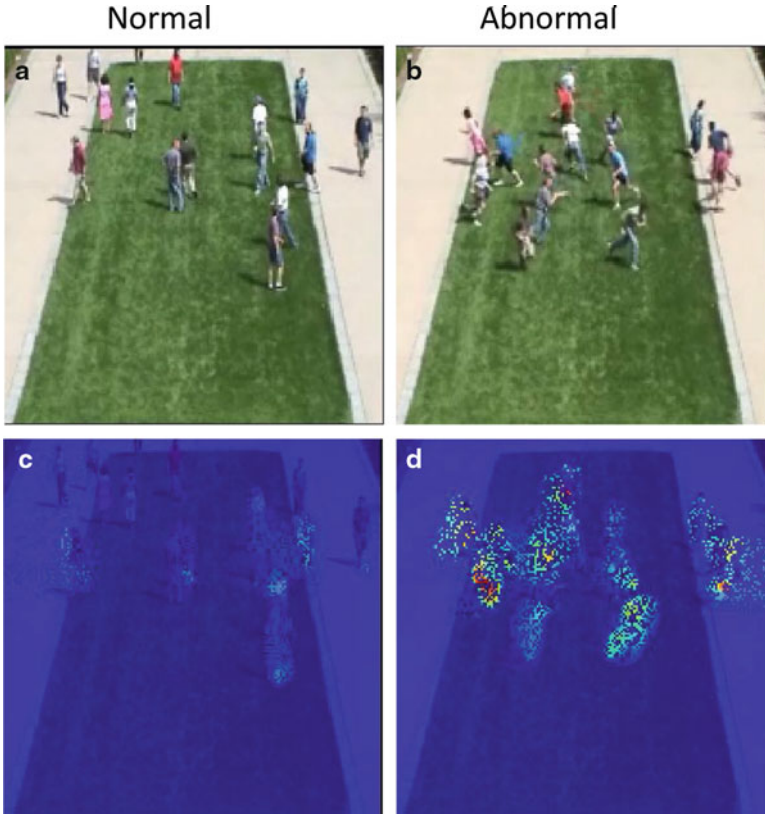
Fig. 15.3 An illustration of the proposed scheme. (**a**) Input normal frame. (**b**) Interaction force corresponding to (**a**). (**c**) Input anomaly frame. (**d**) Interaction force corresponding to (**c**)

the comparison process, like all the other previous algorithms [21]. In practice, we carry out the following steps to decide whether a given frame contains an anomaly or not:

1. First, compute the sum of the interaction forces of a reference frame F_r. This reference frame(s) represents a normal behavior scene in the given video sequence. Actually, all the public datasets considered have an initial (variable, but at least one frame) set of frames representing a normal behavior which can be used as a reference. If k is the number of particles (currently, $k = 15{,}000$), we obtain F_r as follows:

$$F_r = \sum_{i=1}^{k} F_{int}(x_i^{new})|_r \qquad (15.10)$$
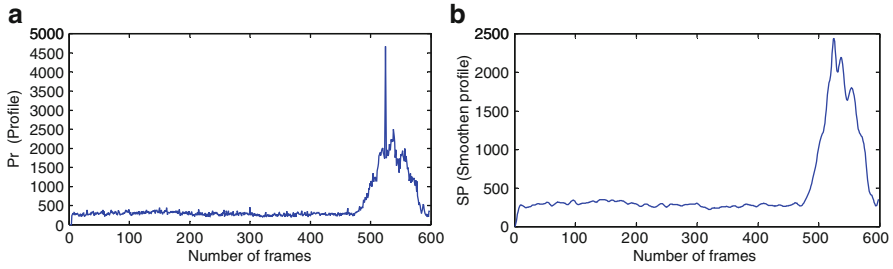
Fig. 15.4 Profile (**a**) before smoothing (**b**) after smoothing

2. Compute the sum of the interaction forces corresponding to all the particles in the current frame F_t as:

$$F_t = \sum_{i=1}^{k} F_{int}(x_i^{new})|_t \tag{15.11}$$

3. Compute the change in the magnitude force at each frame t as:

$$C_t = |F_t - F_r| \tag{15.12}$$

4. Repeat steps 2–3 for all the frames to obtain the profile (values of C_t for all the video frames) corresponding to the change of the force magnitude.

   As an example, Fig. 15.4a shows the profile obtained from a sequence of the UMN dataset after following the above mentioned steps 1–4.

5. Finally, we use the moving average filter to smooth out the short term fluctuations that are present in the obtained profile at the previous step, so to get a smoothed profile C_t^s (see Fig. 15.4b). The moving average is obtained by the simple mean of a few temporally adjacent frames. Once C_t^s is computed, each frame is then classified as either normal or abnormal according to a threshold as follows:

$$L_t = \begin{cases} Abnormal \text{ if } C_t^s > th \\ Normal \text{ otherwise} \end{cases}$$

where C_t^s represents the smoothed profile, th is a threshold value, and L_t holds the final detection result of the given video sequence.

15.3.5 Local Anomaly Detection Scheme Using PSO-SFM

While in the previous section we showed how a frame is classified as either normal or abnormal, the aim of this section is to show how a finer localization of the
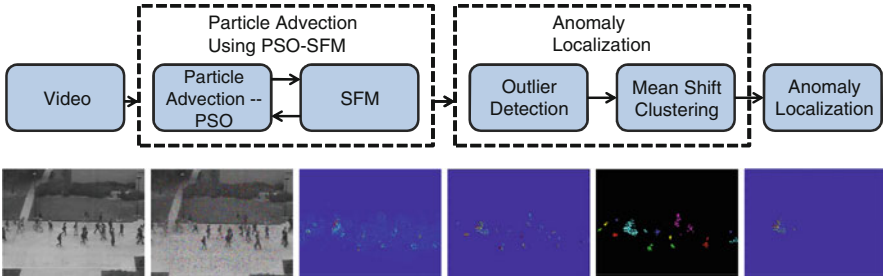
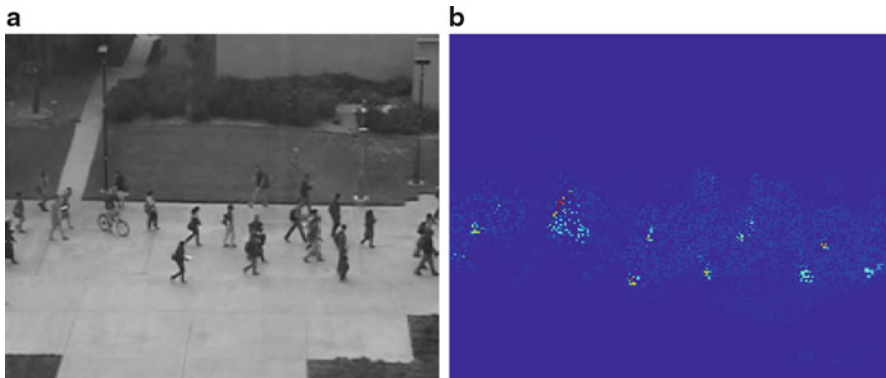**Fig. 15.5** Block diagram of the proposed scheme for anomaly detection and localization

Fig. 15.6 (**a**) Input frame. (**b**) Interaction force

anomaly inside the frame is possible. Figure 15.5 summarizes the proposed scheme for accurate localization of the anomaly in a crowd. The first step is the same interaction force optimization approach presented in Sect. 15.3.3 and used for the global case (see Fig. 15.2).

Figure 15.6a–b show the input frame and the corresponding interaction force, respectively. It is interesting to observe that the highest magnitudes of the force are located in the image regions that move differently from the overall image flow (e.g., the man on the bicycle close to the street lamp). Although patterns of high magnitude of the interaction force over a certain period of time can provide useful information about the presence of an anomaly, not necessarily large magnitudes of the force is a direct consequence of the presence of an anomaly. This is due to the fact that particles are not associated to a whole person, but only to person's parts, so, for instance, legs motion can lead to a high interaction force which is obviously not an anomaly. This motivates us to propose a scheme that can capture the high magnitude patterns over a certain period of time and thereby localize the presence of anomalies in the scene. In order to detect structured interaction forces over time, we use an *outlier detection* scheme to eliminate isolated fluctuations of the social force at each time instant. These "outliers" effects are in general due to the approximation
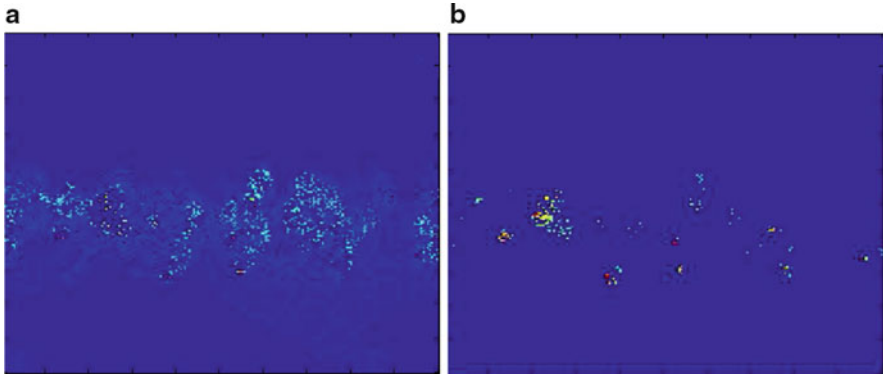
Fig. 15.7 Results of the RANSAC-like algorithm. (**a**) Obtained inliers. (**b**) Corresponding outliers

of the pedestrians velocities with a dense OF computation. For instance, as above observed, we noted that the leg swinging of a walking pedestrian is a cause for false positive (anomaly) detections. This occurs because the local optical flow in this small areas is noisy and may cause some disturbances in the anomaly detection.

The outliers detection process is performed using a custom implementation of the well-known RANdom SAmple Consensus (RANSAC) algorithm [12]. RANSAC is an iterative method used to estimate the parameters of a mathematical model from observed data containing outliers. This algorithm basically assumes that most of the available data consists of inliers whose distribution can be explained by a known parametric model. However, inliers are mixed with outliers which make the direct model parameter estimation inaccurate. Our empirical observations showed that the statistics of the interaction forces associated to a crowd situation in the video datasets can be reasonably well approximated by a Gaussian distribution. Thus, given the interaction force magnitude of the particles at each frame we perform the following steps:

1. Randomly select 5,000 particles (out of 15,000 particles) and their corresponding interaction force magnitude.
2. Estimate the Gaussian distribution using the interaction force magnitude associated with only the selected particles. Let the estimated mean and standard deviation be $\hat{\mu}$ and $\hat{\sigma}$.
3. Consider the remaining particles and evaluate those that are inliers and outliers. Inliers are detected by checking if the particle's force is within the typical $3\hat{\sigma}$ of the estimated model, particles whose force is outside this interval are considered outliers.
4. Repeat the steps 1–3 for R number of iterations, $R = 1,000$ iterations in our case.
5. Finally, choose the Gaussian model with the highest number of inliers.

Figure 15.7a–b show the inliers and outliers obtained using the RANSAC-like algorithm. It is interesting to observe that all high magnitude interaction forces are detected as outliers. In order to achieve a better localization, we perform
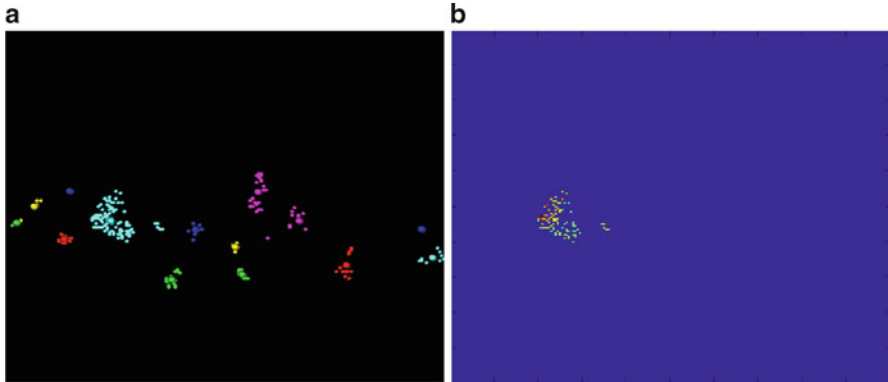
Fig. 15.8 Results of mean-shift clustering. (**a**) Clusters. (**b**) Force magnitude of the largest cluster's particles

Fig. 15.9 An anomaly moving person localized using the positions of the particles in the largest outlier cluster

a spatial clustering of the detected outliers using mean-shift [7, 8] as it works independently on the assumptions regarding the shape of the distribution and the number of modes/clusters. In the end, we finally select the clusters with a number of members larger than a certain threshold, discarding clusters having a small number of particles. This threshold is fixed and kept constant in all the performed experiments; further, assuming that the geometry of the scene is roughly known, this threshold can be set to define the minimal (abnormal) event to be detected.

Figure 15.8a–b show the results of mean-shift clustering and the final anomaly localization obtained after selecting the largest cluster. The positions of the particles of this cluster are plotted on the original input frame in Fig. 15.9. These particles correspond to a moving person on a bicycle, who has been correctly detected as an anomaly because his/her movement does not conform with the movement of the surrounding pedestrians.

15.4 Experiments

In this section we present and discuss the experimental results obtained using the proposed schemes for global and local anomaly detection. We first discuss the results using the global approach and then the experiments performed using the local anomaly scheme.

15.4.1 Experimental Results and Discussion on the Global Anomaly Scheme

To validate the performance of the proposed approach for global anomaly detection, we conducted an extensive set of experiments on four different datasets: UMN [28], PETS 2009 [24], UCF [21], and prison riot dataset (collected by us from the web). In the following experiments, all the video frames are resized to a fixed resolution of 200×200 pixels. For the particle advection scheme, the particle density (i.e., the number of particles) is kept constant at 25 % of number of pixels, and the number of iterations is fixed to 100. To detect the changes of the interaction force magnitude, we use the first frame as the reference frame. This is because in all the datasets the initial (roughly) 40 % of the video frames represents the normal behavior which is then followed by the abnormal behavioral frames. Finally, the performance is validated by plotting the ROC curves obtained over all possible values of the threshold th.

15.4.1.1 UMN Dataset

The UMN dataset consists of 11 video sequences acquired in three different crowded scenarios including both indoor and outdoor scenes. All these sequences exhibit an escape panic scenario: they start with the normal behavior frames followed by the abnormal activity. Figure 15.10 illustrates the results of the proposed scheme obtained on the UMN dataset. Figure 15.10a shows two examples of normal and abnormal crowd behavior frames, respectively, and Fig. 15.10b indicates the corresponding interaction force obtained using the proposed PSO-SFM based particle advection approach. From this figure, it can be observed that the presence of high magnitude of the majority of the particles' interaction force is an evidence that an abnormal frame has occurred. Figure 15.10c shows the detection results of the normal and abnormal frames using step 5 of the global anomaly detection algorithm presented in Sect. 15.3.4. Figure 15.11 shows the detection results obtained on two different sequences of the same UMN dataset. Abnormal frames always correspond to a higher interaction force of the particles.

Figures 15.12 and 15.13 show the performance of the proposed scheme on three different scenes of UMN and on the whole dataset, respectively. The quantitative results in Table 15.1 indicate that the proposed scheme obtained the best performance over different available state-of-the-art methods.
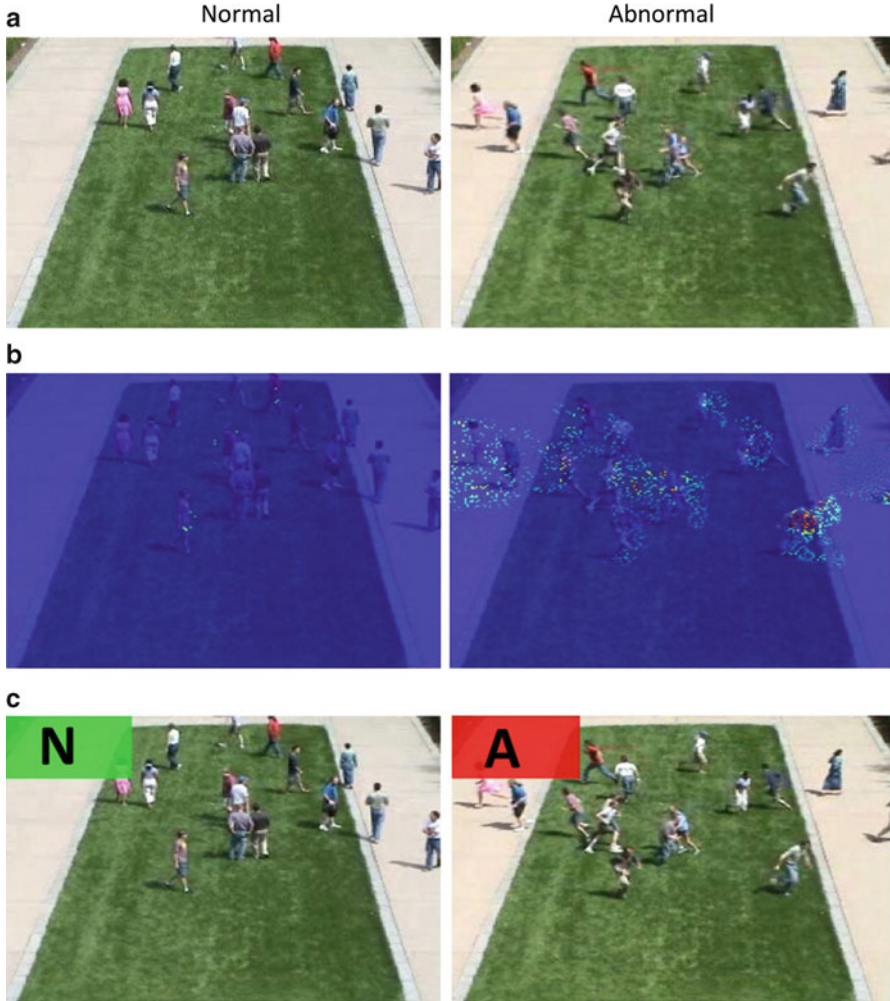
Fig. 15.10 Results on the UMN dataset. (**a**) Input frame. (**b**) Force field. (**c**) Detection (*N* indicates normal and *A* indicates abnormal frame)

15.4.1.2 Prison Riot Dataset

In order to evaluate the proposed method on real applications, we collected a set of real videos from websites such as YouTube and ThoughtEquity.com. The collected video dataset is composed of seven sequences representing riots in prisons that are captured with different angles, resolutions, background and includes abnormality like fighting with each other, clashing, etc. All the collected sequences start with the normal behavior which is then followed by a sequence of abnormal behavior frames. Figure 15.14 shows the interaction force obtained on some of the frames
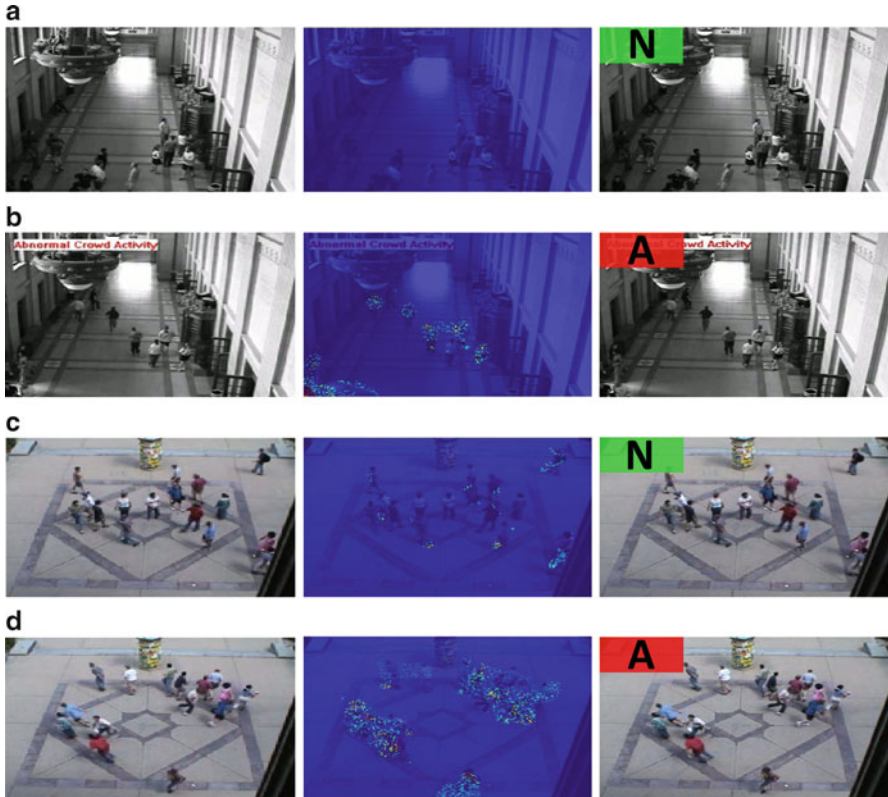
Fig. 15.11 Results of the proposed scheme on other sequences of the UMN dataset. (**a**) Normal behavior in scene 2 with its corresponding interaction force and detection. (**b**) Abnormal behavior in scene 2 with its corresponding interaction force and detection. (**c**) Normal behavior in scene 3 with its corresponding interaction force and detection. (**d**) Abnormal behavior in scene 3 with its corresponding interaction force and detection

of this dataset. Figure 15.15 illustrates the performance of the proposed method on some frames taken from different sequences in this datasets. The ROC curves in Fig. 15.16 demonstrate that the proposed method outperforms the optical flow-based method in distinguishing the abnormal sequences from the normal ones. The quantitative results of this comparison are reported in Table 15.2.

15.4.2   *Results on PETS* 2009 *Dataset*

This section describes the results obtained on PETS 2009 'S3' dataset. This dataset is different from the other datasets used in this chapter, in the sense that abnormality begins smoothly and this makes the detection more challenging because of the
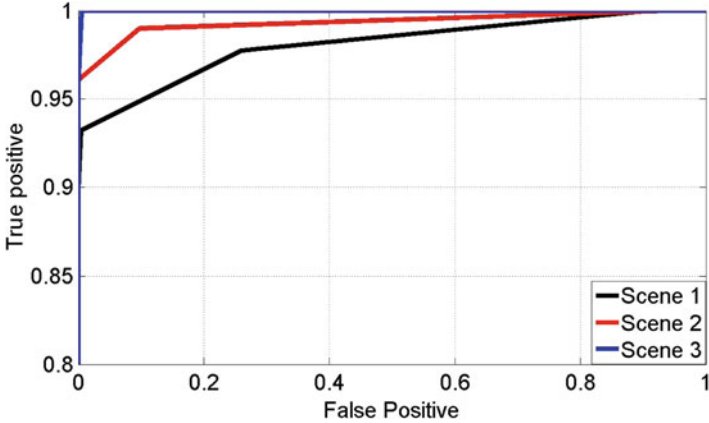
Fig. 15.12 ROC curves of abnormal behavior detection on different scenes in UMN dataset
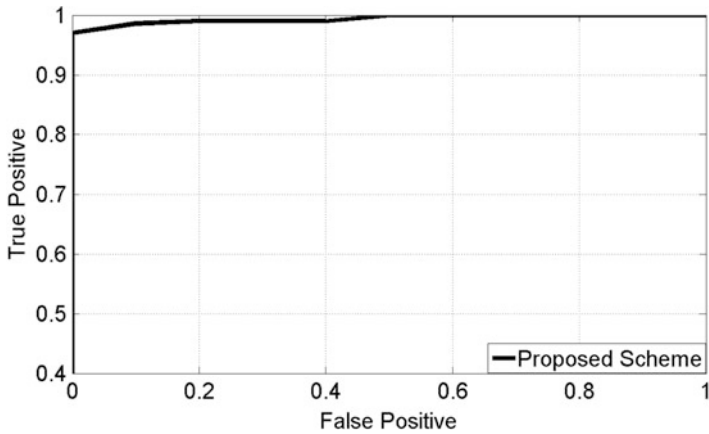

Fig. 15.13 ROC performance on UMN dataset

gradual transaction from normal to abnormal activity. Figure 15.17 shows the interaction force estimated using the proposed scheme on PETS 2009 and Fig. 15.18 shows the corresponding ROC curve. Table 15.3 shows the quantitative results of the comparison, illustrating that the proposed scheme outperforms the optical flow method also with this benchmark.

15.4.2.1   UCF Dataset

Finally, the effectiveness of the proposed algorithm is also evaluated on the UCF dataset [21] composed of 12 video sequences representing normal and abnormal scenes collected from the web. Also in this case, Fig. 15.19 demonstrates that

Table 15.1 Performance of the proposed scheme on the UMN dataset

Method	Area under ROC
Optical flow [21]	0.84
Social force [21]	0.96
Chaotic invariants [30]	0.99
NN [9]	0.93
Sparse reconstruction (scene 1) [9]	0.995
Sparse reconstruction (scene 2) [9]	0.975
Sparse reconstruction (scene 3) [9]	0.964
Sparse reconstruction (full dataset) [9]	0.978
Proposed scheme (scene 1)	**0.9961**
Proposed scheme (scene 2)	**0.9932**
Proposed scheme (scene 3)	**0.9991**
Proposed scheme (full dataset)	**0.9961**

the proposed scheme outperforms the optical flow procedure, and this is further corroborated by the quantitative results reported in Table 15.4 and the qualitative results reported in Fig. 15.20.

The experiments illustrated so far show that the proposed global anomaly detection strategy outperforms the available state-of-the-art methods on realistic datasets like UCF and Prison Riots, other than UMN and PETS 2009 benchmark datasets. The next section is dedicated to testing the local strategy proposed in Sect. 15.3.5.

15.4.3 Experimental Results and Discussion on the Local Anomaly Scheme

To evaluate the performances of the local anomaly detection scheme and compare it with state-of-the-art approaches, we consider two standard datasets used for abnormal activities detection: UCSD [20] and MALL [1] datasets.

15.4.3.1 UCSD Dataset

The UCSD dataset contains two different sets of surveillance videos called PED1 and PED2. The dataset has a reasonable density of people and anomalies including bikes, skaters, motor vehicles crossing the scenes. The PED1 has 34 training and 36 testing image sequence and PED2 has 16 training and 12 test image sequences. These video sequences have two evaluation protocols as presented in [20], namely: (1) frame-level anomaly detection, and (2) pixel-level anomaly detection. At frame-level, we verify if the current frame contains a labeled abnormal pixel. In such a case, the frame is considered containing an abnormal event and compared with the annotated ground truth status (either normal or abnormal). At pixel-level, the
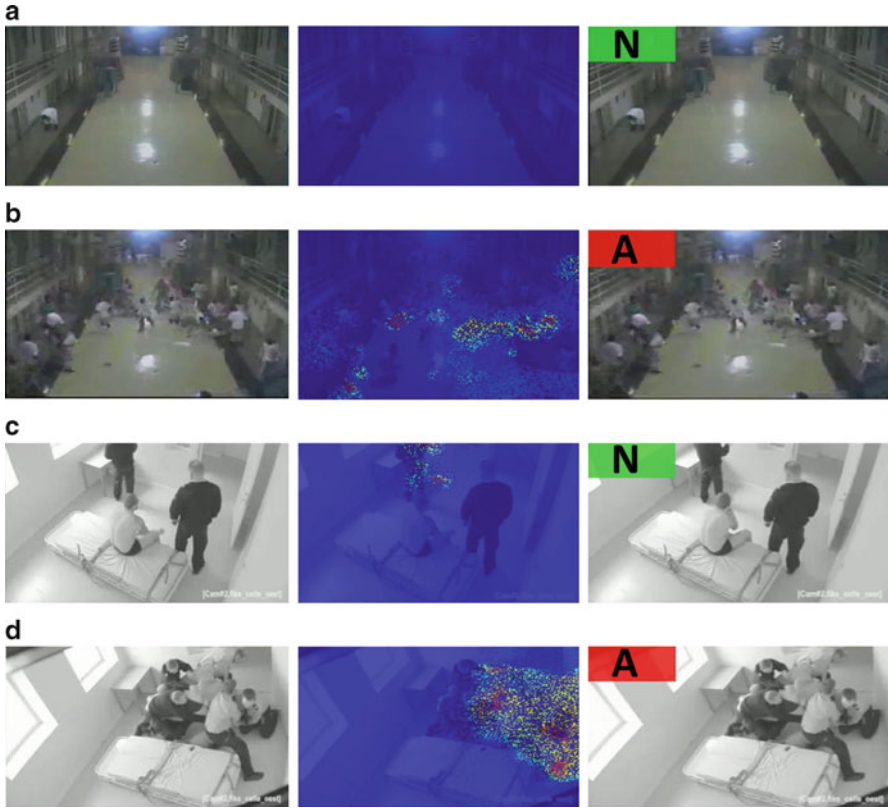
Fig. 15.14 Results of the proposed scheme on the prison dataset. (**a**) A normal behavior frame and its corresponding interaction force and detection result on video sequence 1. (**b**) An abnormal behavior frame and its corresponding interaction force and detection result on sequence 1. (**c**) A normal behavior frame and its corresponding interaction force and detection result on sequence 2. (**d**) An abnormal behavior frame and its corresponding interaction force and detection result on sequence 2

detection of abnormality is compared against the ground truth on a subset of 10 test sequences. If at least 40 % of the detected abnormal pixels match the ground truth pixels, it is presumed that anomaly has been localized otherwise it is treated as a false positive.

Figure 15.21 shows the ROC curve of our method for the frame-level anomaly detection criteria for PED1 and PED2 datasets. We then compare the performance against the state-of-the-art approaches such as the SFM based method [21], MPPCA [16], Adam et al. [1] and Mixture of dynamic textures (MDT) [20]. Table 15.5 shows the quantitative results of the proposed method on frame-level anomaly detection on PED1 and PED2 datasets and Table 15.6 shows the results on anomaly localization. The Equal Error Rate (EER) in Tables 15.5 and 15.6 is defined as the point where false positive rate is equal to false negative rate. Remarkably, the
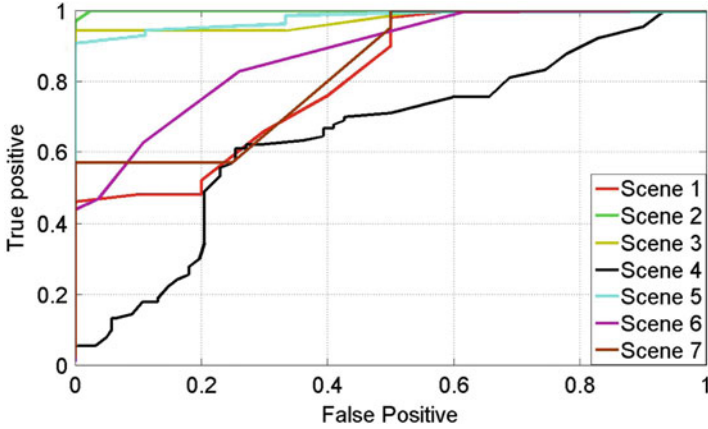
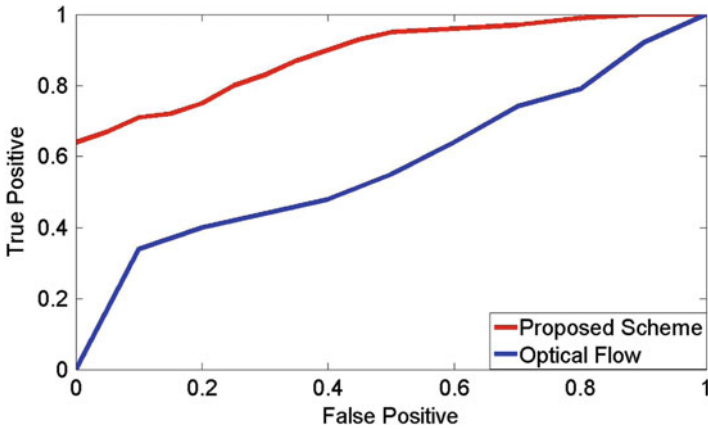**Fig. 15.15** ROC curve of abnormal behavior detection in the different sequences of the prison dataset

Fig. 15.16 ROC curves showing the comparison of the proposed scheme over the optical flow method on the prison dataset

Table 15.2 Performance of the proposed scheme on the prison dataset

Method	Area under ROC
Optical flow	0.5801
Proposed scheme (full dataset)	**0.8903**

proposed method outperforms all the previous approaches on both frame-level and pixel-level detection, reaching the best performances in the frame-level anomaly detection on the PED2 dataset.

Figure 15.22 shows a few frame samples with anomaly detection and localization for the PED1 and PED2 datasets. It can be observed that the proposed method is capable of detecting anomalies even in the far end of the scene (see Fig. 15.22a, last two frames).
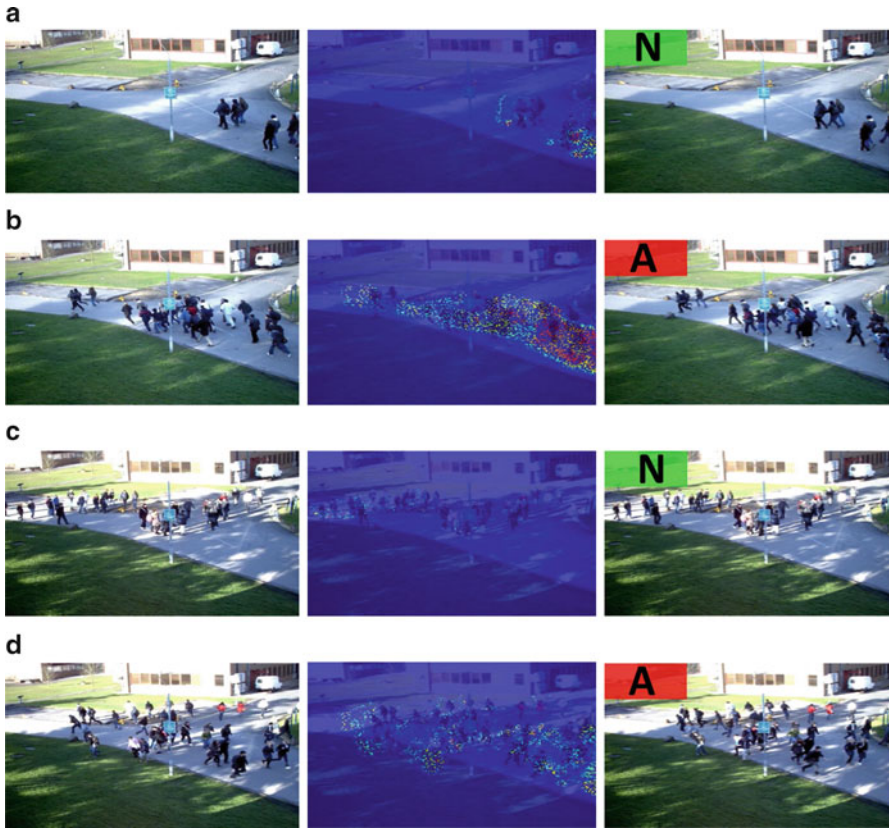
Fig. 15.17 Sample frames from the PETS 2009 dataset (*left column*: input frames, *middle column*: the corresponding interaction force, *right column*: the classification result). (**a**)–(**b**) Sample frames from S3 (14–16). (**c**)–(**d**) sample frames from S3 (14–33)

15.4.3.2 Mall Dataset

The Mall dataset [1] consists of a set of video sequences recorded using three cameras placed in different locations of a shopping mall during working days. The annotated anomalies in such dataset are individuals running erratically in the scene. The evaluation protocol uses only the frame-level anomaly detection criteria. Figure 15.23 shows some frame samples from this dataset in which the anomaly is detected using the proposed method. Table 15.7 shows that the proposed method is extremely accurate in detecting all the frames with an anomaly. Moreover, our approach outperforms the state-of-the-art schemes with respect to the best Rate of Detection (RD) and fewer False Alarm (FA).
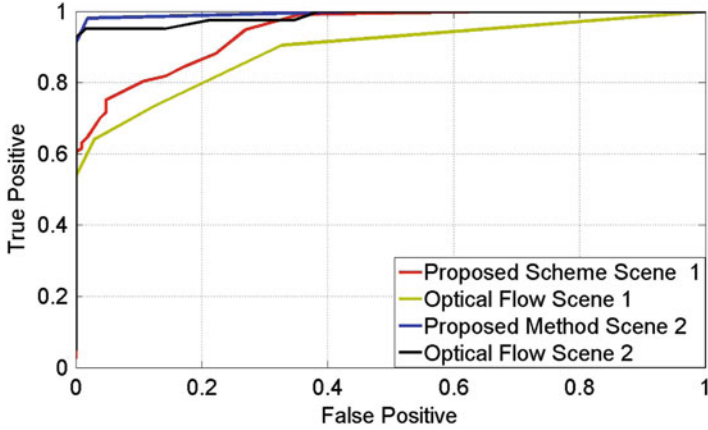
Fig. 15.18 The ROC curves of abnormal behavior detection in the PETS 2009 database

Table 15.3 Performance of the proposed scheme on PETS 2009 dataset

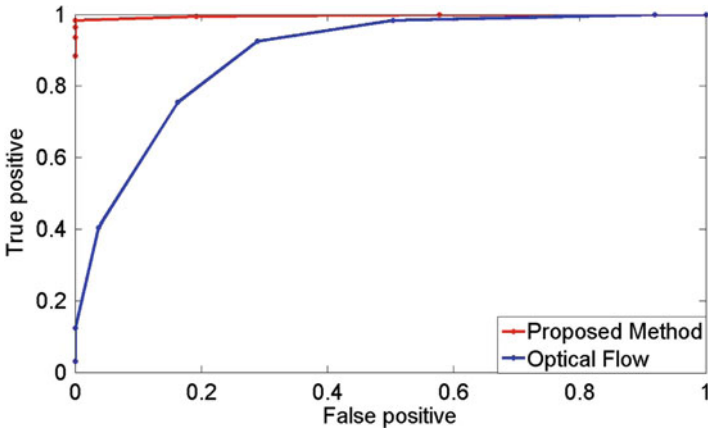| Method	Area under ROC
Optical flow scene 1	0.8834
Proposed scheme scene 1	**0.9414**
Optical flow scene 2	0.9801
Proposed scheme scene 2	**0.9914**

Fig. 15.19 The ROC curves of abnormal behavior detection in the UCF dataset

15.5   Conclusion

We proposed a new particle advection scheme for both global and local anomaly detection in crowded scenes. The main contribution of this work lies in introducing the optimization of the evolving interaction force and performing particle advection

Table 15.4 Performance of the proposed scheme on UCF dataset

Method	Area under ROC
Optical flow	0.884
Proposed scheme (full dataset)	**0.986**

Table 15.5 Equal error rates for frame level anomaly detection on PED1 and PED2 datasets

Approach	SF [21] (%)	MPPCA [16] (%)	Adam et al. [1] (%)	MDT [20] (%)	Proposed method (%)
PED1	31	40	38	25	21
PED2	42	30	42	25	14
Average	37	35	40	25	17

Table 15.6 Anomaly localization: detection rate at the EER

Method	SF [21] (%)	MPPCA [16] (%)	Adam et al. [1] (%)	MDT [20] (%)	Proposed method (%)
Localization	21	18	24	45	**52**

Table 15.7 Performances on the Mall dataset

Dataset	Methods	RD	FA
Mall Cam 1	Adam et al. [1]	95 % (19/20)	1
	Proposed method	100 % (20/20)	2
Mall Cam 2	Adam et al. [1]	100 % (17/17)	6
	Proposed method	100 % (17/17)	4
Mall Cam 3	Adam et al. [1]	95 % (20/21)	4
	Proposed method	100 % (21/21)	3

to capture the optimized interaction force according to the underlying optical flow. The main advantage of the proposed scheme is that the whole anomaly detection/localization process is carried out without any learning phase. This further justifies the applicability of our proposed scheme for real world applications. Finally, empirical results have also indicated that our method is robust and highly performing in detecting abnormal activities on very different types of crowded scenes.

References

1. Adam, A., Rivlin, E., Shimshoni, I., Reinitz, D.: Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 555–560 (2008)
2. Ali, S., Shah, M.: A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–6. Los Alamitos, CA, USA (2007)
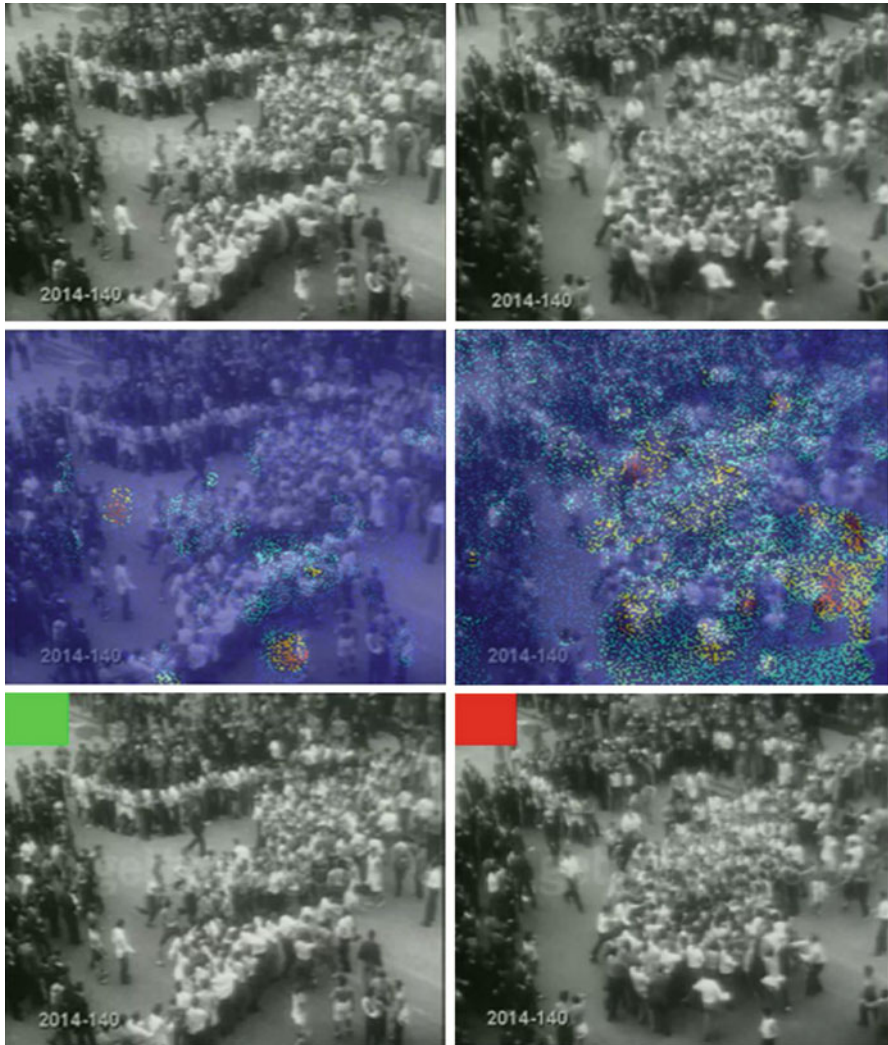
Fig. 15.20 Sample frames from UCF dataset indicating normal (*left*) and abnormal (*right*) frames

3. Antic, B., Ommer, B.: Video parsing for abnormality detection. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 2415–2422. Los Alamitos, CA, USA, (2011)
4. Barnard, K., Duygulu, P., Freitas, D.N., Forsyth, F., Blei, D., Jordan, M.: Matching words and pictures. J. Mach. Learn. Res. **3**(1), 1107–1135 (2003)
5. Blei, M.D., Ng, Y.A., Jordan, I.M.: Latent dirichlet allocation. J. Mach. Learn. Res. **34**(1), 993–1022 (1981)
6. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Proceedings of European Conference on Computer Vision (ECCV), pp. 1–10. Prague (2004)
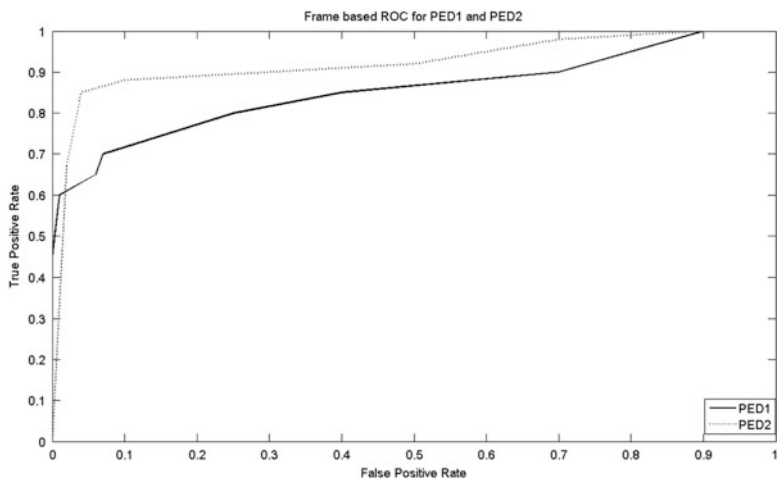
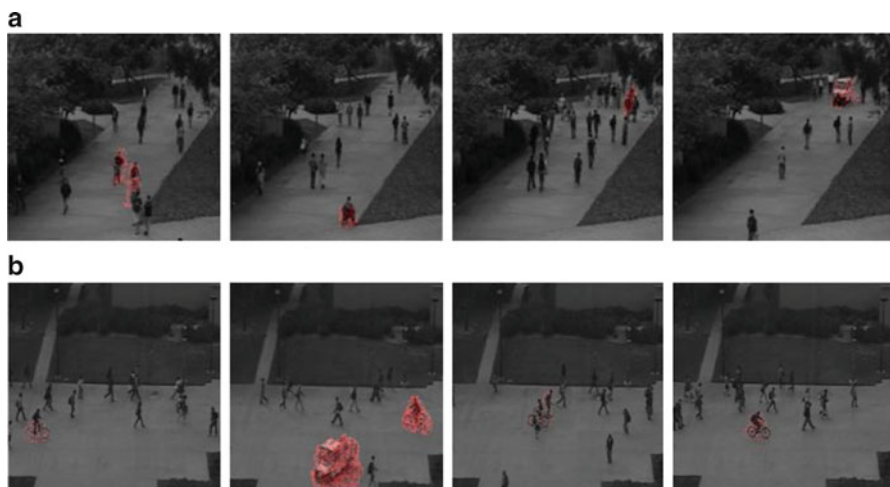**Fig. 15.21** ROC curves obtained on UCSD PED1 and PED2 datasets

Fig. 15.22 (Color online) Examples of anomaly frame detection and localization on PED1 (**a**) and PED2 (**b**) datasets (best viewed in color)

7. Cheng, Y.: Mean shift, mode seeking and clustering. IEEE Trans. PAMI **17**(8), 790–799 (1995)
8. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. PAMI, Colorado Springs, Colorado, USA, **24**(5), 603–619 (2002)
9. Cong, Y., Yuan, J., Liu, J.: Sparse reconstruction cost for abnormal event detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–10. Colorado Springs, Colorado, USA (2011)
10. Cristani, M., Raghavendra, R., Del Bue, A., Murino, V.: Human behavior analysis in video surveillance: a social signal processing perspective. Neurocomputing, **100**, 86–97 (2013)

Fig. 15.23 (Color online) Examples of anomaly detection on the Mall dataset. (**a**) Mall camera 1. (**b**) Mall camera 2. (**c**) Mall camera 3 (best viewed in color)

11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 886–893. San Diego, CA, USA (2005)
12. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(1), 381–395 (1981)
13. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. Phys. Rev. E **51**(4), 42–82 (1995)
14. Jacques, J.C.S., Jr., Raupp Musse, S., Jung, C.R.: Crowd analysis using computer vision techniques: a survey. IEEE Signal Process. Mag. **27**(5), 66–77 (2010)
15. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, pp. 1942–1948. Washington, DC, USA (1995)
16. Kim, J., Grauman, K.: Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921–2928. Miami, Florida, USA (2009)
17. Kratz, L., Nishino, K.: Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1446–1453. Miami, Florida, USA (2009)

18. Krausz, B., Bauckhage, C.: Automatic detection of dangerous motion behavior in human crowds. In: Proceedings of IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), pp. 224–229. Washington, DC, USA (2011)
19. Lekien, F., Marsden, J.: Tricubic interpolation in three dimensions. J. Numer. Methods Eng. **63**(3), 455–471 (2005)
20. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1975–1981. San Francisco (2010)
21. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 935–942. Miami, Florida, USA (2009)
22. Mehran, R., Moore, B., Shah, M.: A streakline representation of flow in crowded scenes. In: Proceedings of European Conference on Computer Vision (ECCV), pp. 1–10. Heraklion, Crete, Greece (2010)
23. Moore, B., Ali, S., Mehran, R., Shah, M.: Visual crowd surveillance through a hydrodynamics lens. Commun. ACM **54**(12), 64–73 (2011)
24. PETS 2009 dataset. http://ftp.cs.rdg.ac.uk/PETS2009/
25. Raghavendra, R., Del Bue, A., Cristani, M., Murino, V.: Abnormal crowd behavior detection by social force optimization. In: Proceedings of Human Behavior Understanding (HBU-2011), pp. 134–145. Amsterdam, The Netherlands (2011)
26. Raghavendra, R., Del Bue, A., Cristani, M., Murino, V.: Optimizing interaction force for global anomaly detection in crowded scenes. In: Proceedings of IEEE Workshop on Modeling, Simulation and Visual Analysis of Large Crowds (MSVLC-2011), pp. 136–143. Barcelona, Spain (2011)
27. Reicher, S.: The Blackwell Handbook of Social Psychology: Group Processes. Blackwell, Oxford (2001)
28. UMN dataset. http://www.mha.cs.umn.edu/movies/crowd-activity-all.avi
29. Wang, X., Ma, X., Grimson, W.E.L.: Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. IEEE Trans. Pattern Anal. Mach. Intell. **31**(3), 539–555 (2009)
30. Wu, S., Moore, B.E., Shah, M.: Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1–6. San Francisco, CA, USA (2010)
31. Zhan, B., Monekosso, D., Remagnino, P., Velastin, S.A., Xu, L.Q.: Crowd analysis: a survey. Mach. Vis. Appl. **19**(5–6), 345–357 (2008)