J Multimodal User Interfaces (2016) 10:307–317
DOI 10.1007/s12193-015-0179-2

ORIGINAL PAPER

Spatial and temporal variations of feature tracks for crowd behavior analysis

Hajer Fradi[1] · **Jean-Luc Dugelay**[1]

Received: 1 October 2014 / Accepted: 18 May 2015 / Published online: 30 May 2015

Abstract The study of crowd behavior in public areas or during some public events is receiving a lot of attention in security community to detect potential risk and to prevent overcrowd. In this paper, we propose a novel approach for change detection, event recognition and characterization in human crowds. It consists of modeling time-varying dynamics of the crowd using local features. It also involves a feature tracking step which allows excluding feature points on the background and extracting long-term trajectories. This process is favourable for the later crowd event detection and recognition since the influence of features irrelevant to the underlying crowd is removed and the tracked features undergo an implicit temporal filtering. These feature tracks are further employed to extract regular motion patterns such as speed and flow direction. In addition, they are used as an observation of a probabilistic crowd function to generate fully automatic crowd density maps. Finally, the variation of these attributes (local density, speed, and flow direction) in time is employed to determine the ongoing crowd behaviors. The experimental results on two different datasets demonstrate the effectiveness of our proposed approach for early detection of crowd change and accurate results for event recognition and characterization.

Keywords Crowd tracking · Local features · Event recognition · Change detection

1 Introduction

There is currently significant interest in visual surveillance systems for crowd analysis. In particular, the study of crowd behavior in public areas or during some public events is receiving a lot of attention for crowd safety to detect potential dangerous situations and to prevent overcrowd (e.g. in religious or sporting events). Many stadium tragedies could illustrate this problem, as well as the Love Parade stampede in Germany and the Water Festival stampede in Colombia. The succession of such deadly accidents emphasizes the need for analyzing crowd behaviors by providing high-level description of the actions and the interactions of and among the objects in crowds. That is an extremely important information for early detection of unusual situations in large scale crowd to ensure assistance and emergency contingency plan.

In this paper, we propose a novel approach to automatically detect abnormal crowd change and to recognize crowd events in video sequences. It is based on analyzing temporal and spatial distributions of persons using long-term trajectories within a sparse feature tracking framework. The idea mainly consists of using low-level local features to represent individuals in the scene. Also, a feature tracking step is involved in the process to alleviate the effects of components irrelevant to the crowd using motion information. By following this strategy, we avert typical problems encountered in detection and tracking of persons in high density crowds, such as dynamic occlusions and extensive clutter.

In addition to the increasing need for automatic detection and recognition of crowd events, our study is motivated by the necessity of implying density estimation in such high level applications since the risk of dangerous events increases when a large number of persons is involved. In the simplest

✉ Hajer Fradi
fradi@eurecom.fr

[1] EURECOM, Campus Sophia Tech 450 route des Chappes, 06410 Biot-Sophia Antipolis, France

forms, the used crowd density measure could be the number of persons [14] or the level of the crowd [16]. However, these measures have the limitation of giving a global information for the entire image and discarding local information about the crowd. We therefore resort to another crowd measure, in which local information at pixel level substitutes a global number of people or a crowd level by frame. The alternative solution [15] is indeed more appropriate as it enables both the detection and the location of potentially crowded areas.

To achieve an improved overall performance, additional information about local density is employed together with regular motion patterns as crowd attributes. These attributes which are first extracted from long-term trajectories, are modeled by histograms to describe the event or the behavior state of a motion crowd. Then, their application to crowd behavior analysis is demonstrated in two steps: First, the temporal stability of these attributes is used for crowd change detection. Second, crowd event recognition is carried out by classifying a feature vector concatenating these histograms. Also, for better video understanding, these attributes are employed to characterize crowd events by providing rich information about their variations in time, the localization of the event, and how many persons participate to a detected event.

The remainder of the paper is organized as follows: The next section revises the state-of-the-art on crowd event detection and recognition. Section 3 presents our sparse feature tracking framework based on extracting long-term trajectories of local features. Details about crowd attributes (local density and motion patterns) are given in Sect. 4. In Sect. 5, we explain how to use these attributes in order to detect crowd change and to recognize crowd events. The application of these attributes to crowd event characterization is presented in Sect. 6. A detailed evaluation of our work follows in Sect. 7. Finally, we briefly conclude and give an outlook of possible future works.

2 Related works

Crowd behavior analysis has recently attracted research attention. This problem covers different subproblems such as crowd change or anomaly detection [5,6,11,18,20,26], and crowd event recognition or characterization [2,8,10,17,19,23,29], in which the goal is to automatically detect changes or to alternatively recognize crowd events in video sequences. In general, there are three main categories of crowd behavior analysis methods. The first category is known as microscopic approaches where the crowd is considered as a collection of individuals who have to be segmented, detected and/or tracked to analyze their crowd behavior. This category includes the Social Force Model [20] which is based on local characteristics of pedestrian motions and interactions, or trajectory-based methods [12,18]. These methods face considerable difficulties to recognize activities inside the crowd because person detection and tracking tasks are affected by occlusions.

In the second category known as macroscopic methods, the crowd is treated as a whole and a global entity for analysis [6,8,9]. For this purpose, scene modeling techniques are used to capture the main features of the crowd behavior. These methods focus on modeling group behaviors instead of determining the motion of individuals which makes them less complex compared to microscopic methods. Hence, they could be applied to analyze scenes of medium to high crowd density. The third category known as hybrid methods studies the crowd at a microscopic and macroscopic levels. They inherit both properties to handle the limitations of each category of methods and to complement each others for better performance [2,4,17,28].

Our proposed method is of hybrid nature since it incorporates optical flow information into extracted local features and it examines long-term trajectories to capture both global and local attributes. These attributes have the advantages of capturing the spatial and temporal variations of feature tracks simultaneously. Consequently, they convey rich information about the spatial distributions and mouvements of pedestrians in the scene which are strongly related to the ongoing crowd behaviors.

While most of existing works rely on optical flow information between consecutive frames, in our approach we extend this information to build trajectories in order to accurately represent the motion with the video. Also, the generated feature tracks undergo an implicit temporal filtering step which makes them smoother.

Another substantial contribution of this paper, is the use of local crowd density in addition to the commonly used crowd motion forms (speed and orientation). We consider it as an important cue for early detection of crowd event and it could complement crowd dynamics (motion) information. For example, walking/running events are typically recognized by measuring the speed. However, it is also important to provide additional information about the number or the density of individuals moving at high speed. Other crowd events such as crowd formation/splitting have been analyzed using the direction of optical flow, again this information is not sufficient, because a large number of individuals has to be involved and to participate to crowd formation. Another example that justifies the relevance of using crowd density for event characterization is the blocking situations in large scale crowd, in this case relying on motion information is not enough since there is no enough spaces to move, as a result the speed slows down. These examples illustrate the need to use density as additional cue for characterizing crowd events, also it helps to localize crowded regions.

3 Crowd tracking

Although there are different approaches to the tracking problem, their applications are limited to scenes with few and easily perceptible constituents. Generally, the application of conventional tracking algorithms on videos of high dense crowds is challenging and is encountered by many issues. Actually, crowded scenes exhibit some particular characteristics rendering the problem of multi-target tracking more difficult than in scenes with few people: Firstly, due to the large number of pedestrians within extremely crowded scenes, the size of a target is usually small in crowds. Secondly, the number of pixels of an object decreases with a higher density due to the occlusions caused by inter-object interactions. Thirdly, constant interaction among individuals in the crowd makes it hard to discern them from each others. Finally and as the most difficult problem, full target occlusions that may occur (often for a long time) by other objects in the scene or by other targets.

All the aforementioned factors contribute to the loss of observation of target objects in crowded videos. These challenges are added to the classical difficulties hampering any tracking algorithm such as: changes in the appearance of targets related to the camera view field, the discontinuity of trajectories when the target exits the field of view and re-appears later again, cluttered background, and similar appearance of some objects in the scene. Given all these difficulties, conventional human detection or tracking paradigms fail in such cases. To overcome this problem, alternative solutions which consist of tracking particles [19,20,29] or local features [3,11,18,23] instead of pedestrians have been proposed. Other methods operate on foreground masks and consider them as the regions of interest [5,6,9], called activity area in [6].

In this paper, our proposed approach for crowd tracking is based on tracking local features. First, to infer the contents of each frame under analysis we extract local features. Then, we perform local features tracking using the Robust Local Optical Flow algorithm from [25] and a point rejection step using forward-backward projection. An illustration of the crowd tracking modules is shown in Fig. 1. The remainder of this section describes each of these system components.

3.1 Extraction of local features

One of the key aspects of crowd tracking is feature extraction. Under the assumption that regions of low density crowd tend to present less dense local features compared to high-density crowd, we propose to use local features as a description of the crowd by relating dense or sparse local features to the crowd size. For local features, we assess features from accelerated segment test (FAST) [22].

Features from accelerated segment test is proposed for corner detection in a fast and a reliable way. It depends on a wedge model style corner detection. Also, it uses machine learning techniques to find automatically optimal segment test heuristics. The segment test criterion considers 16 surrounding pixels of each corner candidate P (of intensity I_P). Then, P is labeled as corner if there exist n contiguous pixels in the circle (of 16 pixels) that are all brighter than $(I_P + t)$ or all darker than $(I_P - t)$. In the experiments, n and t are set to 12 and 30, accordingly.

The reason behind selecting this feature for crowd measurement is as follows: FAST was proposed for corner detection in a reliable way. It has the advantage of being able to find small regions which are outstandingly different from their surrounding pixels. In addition, FAST was used in [7] to detect dense crowds from aerial images and the derived results demonstrate a reliable detection of crowded regions.

3.2 Local features tracking

Local features tracking is performed by assigning motion information to the detected features. In our framework, we

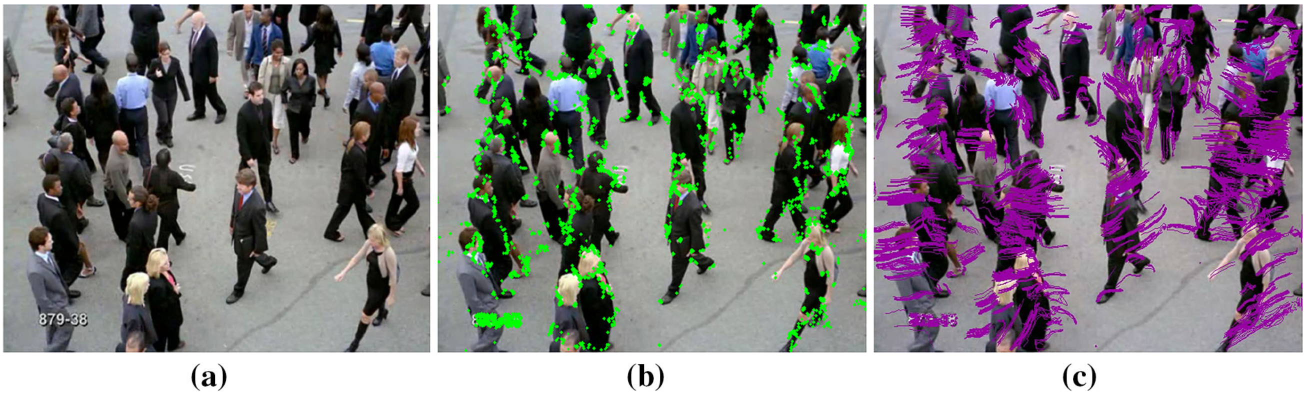

Fig. 1 Illustration of the proposed crowd tracking using local features: **a** exemplary frame, **b** local features extraction using FAST, **c** feature tracks over time using RLOF

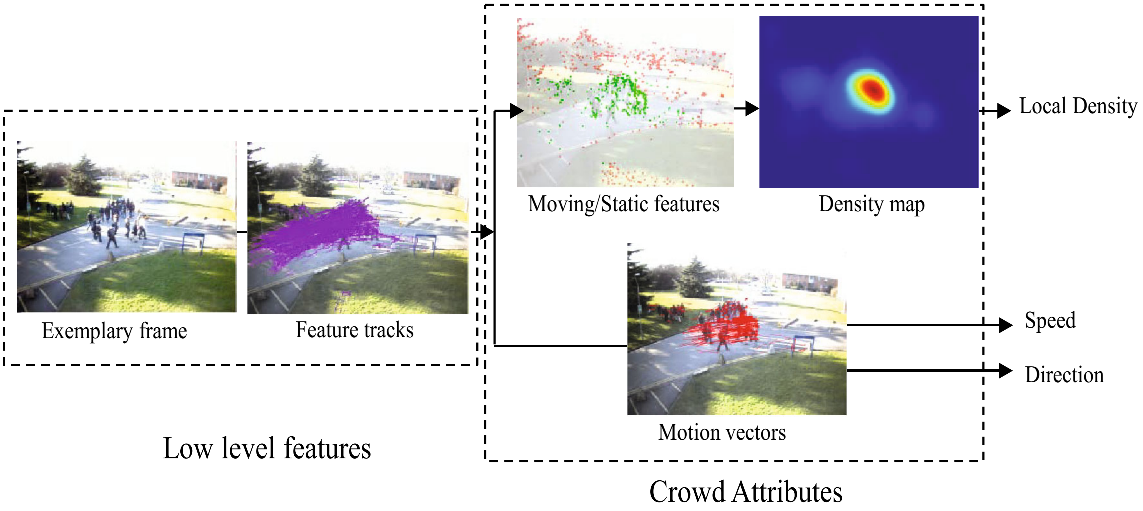

Fig. 2 Illustration of the proposed crowd attributes: crowd tracking using local features, estimation of crowd density map after distinction between moving (*green*) and static (*red*) features, estimation of speed and flow direction from motion vectors

apply the robust local optical flow (RLOF) [24,25], which computes accurate sparse motion fields by means of a robust norm.[1] A common problem in local optical flow estimation is the choice of feature points to be tracked. Depending on texture and local gradient information, these points often do not lie on the center of an object but rather at its borders and can thus be easily affected by other motion patterns or by occlusion. While RLOF handles these noise effects better than the standard Kanade-Lucas-Tomasi (KLT) feature tracker [27], it is still not prone against all errors. This is why we establish a forward-backward verification scheme where the resulting position of a point is used as input to the same motion estimation step from the second frame into the first one. Points for which this "reverse motion" does not result in their respective initial position are discarded. For all other points, motion information is aggregated to form trajectories by connecting motion vectors computed on consecutive frames. This results a set $\mathcal{T}_k$ of n_k trajectories in every time step k:

$$\begin{aligned} \mathcal{T}_k &= \left\{T_1^k, \ldots, T_{n_k}^k \right| \\ T_i^k &= \left\{X_i\left(k - \Delta t_i^k\right), Y_i\left(k - \Delta t_i^k\right), \ldots, X_i(k), Y_i(k)\right\}\right\} \end{aligned} \quad (1)$$

where Δt_i^k denotes temporal interval between the start and the current frames of a trajectory T_i^k. $\left(X_i\left(k-\Delta t_i^k\right), Y_i\left(k-\Delta t_i^k\right)\right)$, and $(X_i(k), Y_i(k))$ are the coordinates of the feature point at its start and current frames, respectively. The advantage of using trajectories in our system instead of computing the motion vectors only between two consecutive frames is that outliers can be filtered out and the overall motion information is more reliable and less affected by noise, more details about processing these trajectories are presented in the next section.

4 Crowd event attributes

For crowd event attributes, we simultaneously consider local density, speed and orientation. These attributes are extracted from our proposed sparse feature tracking framework described in Sect. 3. For local density, a probability density function (pdf) on the positions of moving local features using a Gaussian kernel density is computed, whereas, speed and orientation are estimated from motion vectors. An illustration of the modules of crowd attributes extraction is shown in Fig. 2.

4.1 Local crowd density

Our proposed local crowd density is estimated by measuring how close local features are. This is based on the observation the more local features come towards each other, the higher crowd density is perceived. Since the extracted local features defined in Sect. 3.1 contain components irrelevant to the crowd density, we need to add a separation step between foreground and background entities to our system. This feature selection process can be optimally done by computing the overall motion Γ_i^k of each trajectory T_i^k (tracks of an extracted local feature). Γ_i^k, which denotes the average displacement between $\left(k - \Delta t_i^k\right)$th and the current frame k, is compared to a small constant ζ (set to 1). Moving features are then identified by the relation $\Gamma_i^k > \zeta$ while others are considered as part of static background.

[1] http://www.nue.tu-berlin.de/menue/forschung/projekte/rlof.

After filtering out static features, the crowd density map is defined as a kernel density estimate based on the positions of moving local features. For a given video sequence of N frames $\{I_1, I_2, \ldots, I_N\}$, if we consider a set of m_k moving local features extracted from a frame I_k at their respective locations $\{(x_i, y_i), 1 \leq i \leq m_k\}$, the corresponding density map C_k is defined at a pixel position (x, y) as follows:

$$C_k(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^{m_k} \exp - \left(\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma^2} \right) \tag{2}$$

where σ is the bandwidth of the 2D Gaussian kernel which defines the effect of each local feature on the density calculation. σ has to be large enough to guarantee the involvement of local features which are close to (x, y) in the calculation of the density at this position. However, there are some rules to properly choose σ for different situations within the same video and for different videos as well. First, within the same video, σ is adaptively set according to a perspective map, in order to deal with the effects of perspective distortions. For this problem, the distance between (x, y) and the other local features has to be updated as well by the same criteria. Second, σ has to be also proportional to the resolution of the video. This setting strategy of σ guarantees its invariance to scale and to resolution changes.

The resulting crowd density map characterizes the spatial variation of the crowd thanks to the probability density function involved in the process. This spatial variation that arises across the frame conveys rich information about the distributions of pedestrians in the scene.

4.2 Crowd motion: speed and orientation

The feature tracks defined in Sect. 3 are first used to show the spatial distributions of the crowd by estimating crowd density maps based on the positions of moving local features. Second, they are used to extract crowd motion information. It proceeds as follows: after filtering out static features (of zero trajectory lengths because they are stationary along frames, or of small trajectory lengths because of the noise in video acquisition, or dynamic background), for the remaining local features, we restrict the history of each 2D trajectory over last few frames (set experimentally to 50 frames) because otherwise by considering the whole trajectory an augmentation in the speed will not be detected early, also the flow direction could be less precise. Then, the overall motion Γ_i^k of a trajectory T_i^k is compared to a certain threshold β which is empirically set to 1/3 of average motion at each frame k. The trajectory is considered for further processing only if $\Gamma_i^k > \beta$, while other short-term trajectories of small length (occur because of tiny movement of crowd) are filtered out to not affect the computation of speed and orientation.

Once the set of useful trajectories is determined, we compute the speed as the quotient of the trajectory length divided by the number of frames being tracked. For flow direction, we consider the orientation of motion vector formed by the start and the current position of each trajectory.

5 Abnormal change detection and event recognition

Overall, the spatio-temporal crowd measures introduced by density maps and motion vectors convey rich information about the spatial distributions and the movements of pedestrians in the scene which are strongly related to their behaviors. For this goal, we first model the crowd attributes by histograms, see Sect. 5.1. Then, the application of these attributes for crowd behavior analysis is demonstrated in two steps: First, the variation of a stability measure (using the histograms) in time is employed to detect change or abnormal event, see Sect. 5.2. Second, a feature vector concatenating these histograms is used for event recognition, see Sect. 5.3.

5.1 Crowd modeling

Each crowd attribute is encoded by 1D-histogram. Given the crowd density map C_k at a frame k, the local density information is quantized into N_d bins. We have chosen $N_d = 5$ according to Polus definition [21] of crowd levels (free, restricted, dense, very dense and jammed flow). Then, to group together motion vectors of the same direction, we quantize the orientation Θ into N_Θ bins. N_Θ is set to 8 bins, which results orientation bin size $\Delta_\Theta = 45$ degrees. As proposed in [5], the speed is quantized into $N_s = 5$ classes: very slow, walking, walking fast, running, and running fast. Since the speed is computed in the image coordinates, its changes can be affected by the perspective distortions, due to the fact that when people are moving away from the camera, their motion vectors are becoming of small lengths. That is why, we need to rectify these effects on the speed. To achieve this goal, we weight the lengths of trajectories according to a perspective map, which is approximated by linearly interpolating the perceived height of a reference person in the two extreme lines of the scene [14].

5.2 Crowd change detection

According to the procedure described so far, at each frame k, we obtain three histograms $H_d(k)$, $H_\Theta(k)$, and $H_s(k)$ which denote, respectively, the histograms of density, orientation, and speed. If the motion patterns and the density of the crowd remain similar within a period of time, the corresponding histograms are similar as well. Whereas, if a change occurs in the crowd behavior, that would generate dissimilarities between the histograms.

For histogram comparison in time, we adapt the same strategy as in [5]: we compare the density and the motion patterns at each frame with the those of a set of previous frames. For each histogram $H_i(k)$ at time k, a similarity vector $S_i(k)$ is defined as:

$$S_i(k) = (C(H_i(k), H_i(k-\Delta t_1)), \\ C(H_i(k), H_i(k-\Delta t_2)), \ldots, C(H_i(k), H_i(k-\Delta t_n))) \tag{3}$$

n is the number of frames used in the comparison (n is experimentally set to 25), Δt_j are the frame steps, and C is the histogram correlation defined between H_1 and H_2 as:

$$C(H_1, H_2) = \frac{\sum_p \left(H_1(p) - \overline{H_1}\right)\left(H_2(p) - \overline{H_2}\right)}{\sqrt{\sum_p \left(H_1(p) - \overline{H_1}\right)^2 \sum_p \left(H_2(p) - \overline{H_2}\right)^2}} \tag{4}$$

where $\overline{H}$ is the mean value of H, and p is the histogram bin.

Similar to [5], we define the temporal stability $\sigma_i(k)$ of each histogram $H_i(k)$ as the weighted average of $S_i(k)$:

$$\sigma_i(k) = \omega^T S_i(k), \\ \omega = \frac{1}{\sum_{j=1}^{n} e^{\lambda \Delta t_j}} \left(e^{-\lambda \Delta t_1}, e^{-\lambda \Delta t_2}, \ldots, e^{-\lambda \Delta t_n}\right) \tag{5}$$

λ denotes the decay constant, $\Delta t_j = j\Delta t$ (Δt is a constant). λ and Δt are set to 0.52 and 0.25, respectively.

In our approach, a change is detected if the similarity between the current frame and the previous frames for one of the crowd attributes (local density, speed, or orientation) is low. For this, we compare each temporal stability $\sigma_i(k)$, $1 \leq i \leq 3$ to an adaptive threshold $\tau_i(k)$ computed as the half average of the temporal stability values σ_i between $(k - \Delta t_1)$ and $(k - \Delta t_n)$:

$$\tau_i(k) = \frac{1}{2n} \sum_{j=1}^{n} \sigma_i(k - \Delta t_j) \tag{6}$$

5.3 Event recognition

The proposed crowd attributes are also used to recognize crowd events. In particular, six crowd events are modeled namely, walking, running, evacuation, local dispersion, crowd formation and crowd splitting. In our approach, we propose to perform event recognition by classification. For testing, given a new frame $\mathbf{x}$, we aim at classifying it into one of the events $y^* \in \mathcal{Y}$, which maximizes the conditional probability:

$$y^* = \arg\max_{y \in \mathcal{Y}} P(y|\mathbf{x}, \theta^*) \tag{7}$$

where θ^* are learned from the training data. This can be performed by SVM classification, and for the feature vector, we concatenate the three histograms $H_d(k)$, $H_\Theta(k)$, and $H_s(k)$ into $\mathcal{H}_k$. For classification, we use Chi-Square kernel:

$$K(\mathcal{H}_i, \mathcal{H}_j) = \sum_I \frac{(\mathcal{H}_i(I) - \mathcal{H}_j(I))^2}{\mathcal{H}_i(I) + \mathcal{H}_j(I)} \tag{8}$$

6 Crowd event characterization

We consider that the local density is an important cue to characterize crowd events. In addition, it provides helpful information about the density of people that participate to a detected event, also it is useful to localize the event since it is estimated at local level. The characterization of crowd events is as follows:

6.1 Walking/running

Walking event corresponds to a number of persons moving at low speed. If the speed is high, running event is detected. This can be recognized by computing the average of magnitudes of motion vectors at each frame.

6.2 Evacuation

Evacuation is defined as a sudden dispersion of the crowd in different directions. To recognize this event, direction, speed, and crowd density attributes can be used. This event can be characterized by detecting more than four principal directions which have to be distant from each others. Also, a degradation in the crowd density and an increase in the speed and in the motion area have to be detected to recognize this event.

6.3 Crowd formation/splitting

Crowd formation (or merging) event is recognized when we detect a merge of many individuals coming from different directions towards the same location. For this purpose, distance between main directions can be used. Also, this event is characterized by an increase in the crowd density and a decrease in the motion area. The opposite of crowd formation is crowd splitting event.

6.4 Local dispersion

This event is recognized when people moves locally away from a threat. The same attributes of crowd formation and splitting can be used.

7 Experimental results

7.1 Datasets

To evaluate our proposed approach for crowd change detection, event recognition, and crowd characterization, we use two public datasets: PETS.S3 dataset [13] and the dataset of the University of Minnesota (UMN) [1]. The publicly available UMN dataset has been widely used to distinguish between normal and abnormal crowd activities. This dataset comprises 11 videos in three indoor and outdoor scenes organized as follows: Videos 1:2 belong to scene 1, Videos 3:8 belong to scene 2, and the scene 3 consists of Videos 9:11. Each of these videos can be divided into normal and abnormal parts. Precisely, they illustrate different scenarios of escape event such as crowds running in one direction, or people dispersing from a central point.

For the ground truth, as noticed in some previous works [5,11], the labels of abnormal events shown in the videos are not accurate. There are some time lags in the ground truth labels, for instance in Video1, according to the labels of the ground truth, it is shown that an abnormal event occurs from frame 526, however people started running at frame 484. To overcome this problem, we use the labels of change detection of some videos provided in [5,11], for the other videos we follow the same annotation strategy; we manually label the frame in which the crowd change happens (in particular, in UMN dataset as soon as people start running).

The Section S3. Event Recognition of PETS dataset has been employed to assess crowd event detection and recognition algorithms. This dataset comprises 4 video sequences with the following time-stamps 14:16, 14:27, 14:31 and 14:33. As noticed in [17], some sequences are composed of two video clips, this is the case of 14:16, 14:27, and 14:33, which results seven videos in general. More details about these seven videos are given in Table 1.

These videos depict six classes of crowd events: walking, running, formation (merging), splitting, evacuation, and dispersion. We annotate these videos with the six classes as it is shown in the following Table 2.

Table 1 Videos from PETS

Sequence name	First frame	Last frame
14:16-a	0	107
14:16-b	108	222
14:27-a	0	184
14:27-b	185	333
14:33-a	0	310
14:33-b	311	377
14:31	0	130

S3 used for testing crowd events recognition and characterization algorithms: the first and the last frames of each video sequence

Table 2 The time intervals indicate when a specific event is recognized (from its first frame to the last one)

Events	Video [frames]
Walking	seq.14:16-a [0-40], seq.14:16-b [0-56]
Running	seq.14:16-a [41-107], seq.14:16-b [57-114]
Evacuation	seq.14:33-b [24:66]
Dispersion	seq.14:27-a [96:144], seq.14:27-b [86:134]
Formation	seq.14:33-a [0:180]
Splitting	seq.14:31 [58:130]

Table 3 Comparison of our detection results to the ground truth labels using error frame metric

Seq.	Nb frames	Ground truth	Our Det. changes	e_F
UMN.Video1	625	484	493	0.0144
UMN.Video2	828	665	669	0.0048
UMN.Video3	549	303	319	0.0291
UMN.Video4	685	563	582	0.0277
UMN.Video5	769	492	512	0.0260
UMN.Video6	579	450	466	0.0276
UMN.Video7	895	734	754	0.0223
UMN.Video8	667	454	471	0.0255
UMN.Video9	658	551	551	0
UMN.Video10	677	570	577	0.0103
UMN.Video11	807	717	722	0.0062

7.2 Experiments and analysis

7.2.1 Crowd change detection

For evaluating crowd change detections, accurate detection means early detection as soon as the change occurs. For quantitive evaluation, we employ the relative mean frame error metric proposed in [19]. It is defined as:

$$e_F = N_e/N_{fr} \tag{9}$$

where N_{fr}, N_e denote the total number of frames in the video, and the error frames, respectively, see Table 3.

In this Table, we show the results of change detection for videos from UMN dataset. The comparison of our detection results to the ground truth labels shows satisfactory performances and rather accurate in most videos. In terms of e_F metric (the last column in the Table), the error is small in most cases. In our approach, the delay in the detection of some frames after the event occurs is because of our strategy of detection, in which an abnormal event is detected if the temporal stability is becoming below the dynamic threshold (defined as half the average of temporal stabilities of previous frames). This requires some times to be detected, which

Table 4 Performance of our proposed crowd change detection method in terms of recall and precision using UMN dataset compared to [11, 19, 26]

Approach	Recall (%)	Precision (%)
Proposed approach	92.45	100
AMC approach [11]	94	n/a
STCOG approach [26]	92.28	94.47
Approach in [19]	84.75	100

justifies the delay. At the same time, this strategy is suitable to avoid false alarms.

To demonstrate the effectiveness of our proposed approach, we compare our results to adjacency-matrix based clustering (AMC) method [11], spatial temporal co-occurrence Gaussian mixture models (STCOG) method [26], and to the method proposed in [19], which is based on dense optical flow and particle advection. The precision and recall of all these methods are listed in Table 4. The comparison shows that our method achieves comparable results to [11] in terms of recall. 100 % is achieved in terms of precision which means zero false alarms for all videos, however, the evaluation in terms of precision is not provided for the compared method [11]. For recall we get worse results, but of small margin. The comparison[2] to STCOG method [26] shows better performance for our proposed method. Finally, the results of [19], demonstrate that, similar to our approach, this method succeeds to avoid false alarms, however the delay in the detection is bigger than in our approach.

To conclude, the effectiveness of our proposed approach for crowd change detection has been validated by showing excellent performance in terms of false detections (100 % as precision). For the recall, our approach achieves comparable results regarding the other existing methods. These results are explained by the same reason mentioned before, about the time lags in the detection until the similarity metric becomes less than the dynamic threshold. Also, it is important to mention that UMN dataset does not include events such as crowd formation/splitting, that could justifies that methods based only on motion information (speed and orientation) could achieve satisfactory results.

7.2.2 Crowd event recognition

For crowd event recognition, we use PETS.S3 dataset from view 1 and view 2 in order to increase the available frames. We randomly split this dataset into (75 %) for training and (25 %) for testing. For each test sample, the feature vector using the concatenation of the three histograms is identified as one of the six classes following one-vs-one strategy. We obtain 98.25 % as classification accuracy, when we used the three crowd attributes, and 92.28 % when the histogram of density is not included in the feature vector. These results demonstrate good performance for both cases, which proves the relevance of our proposed crowd tracking framework. A significant improvement (around 6 % in the classification accuracy) using local density as additional crowd attribute is noticed as well. Also, we evaluate the recognition performance with confusion matrix, see Table 5.

As it is shown in this Table, we achieve excellent results for all crowd events including crowd formation/splitting, which justifies again the relevance of our proposed attributes.

For comparisons,[3] we report the classification accuracy on the test set for each class separately, following one-vs-rest strategy, see Table 6. In this table we compare our results to [8], in which the recognition is performed using color, texture and shaped features. Also, we add a comparison to [28], based on Lucas-Kanade optical flow method [27].

For the first compared method, the tests has been done for view 1 and 2 separately. In most cases, we have better results. Also, in [8], some difficulties to recognize the events from view 2 have been reported, which justifies the incapability of this method to deal with different point of view. By comparing our method to [28], we notice that our method has better results in most cases, even though the compared method runs on samples from the same view. This demonstrates that our proposed approach achieves good results independently from the camera point of view.

Overall, these results demonstrate that our proposed approach achieves better performance compared to the other existing methods. These results justify the effectiveness of the proposed sparse feature tracking framework, which accurately represents the motion in the video. In particular, they justify the advantage of using trajectories instead of motion vectors by filtering out outliers and removing noisy information. In addition, the use of crowd density as additional attribute with motion patterns has shown substantial improvement in the classification accuracy. That demonstrates the relevance of this attribute to complement motion information, and consequently to identify crowd behaviors.

7.2.3 Crowd characterization

For evaluating our proposed crowd event characterization, we use PETS.S3 dataset. By following up some measures

[2] These results have to be considered carefully, because in [26] according to the number of frames, we noticed that the authors used one frame out of each three frames. Also, the original ground truths have been used in these results. These two factors may boost the results reported in the compared paper.

[3] Again, these comparisons have to be considered carefully, even though we mostly agree with the compared methods on the ground truth labels, and on the evaluation strategy, we cannot ensure that the algorithms run on the same dataset because of the random selection of training/testing samples.

Table 5 Confusion matrix for event recognition on PETS 2009

	Walking	Running	Splitting	Dispersion	Evacuation	Formation
Walking	97.96	0	0	0	2.04	0
Running	4.17	95.83	0	0	0	0
Splitting	2.78	0	97.22	0	0	0
Dispersion	0	0	0	100	0	0
Evacuation	0	0	0	0	96.88	3.12
Formation	0	0	0	0	0	100

S3 dataset

Table 6 Classification accuracy of our proposed crowd event recognition method compared to Cermeno et al. method [8] and to Xu et al. [28] on test set from PETS.S3 dataset following one-vs-rest strategy

Methods	Views	Walking	Running	Splitting	Dispersion	Evacuation	Formation
Proposed method	View 1 and 2	98.25	99.30	100.00	100.00	98.25	98.60
Cermeno et al. method [8]	View 1	98.87	97.62	98.77	92.16	100	99.25
Cermeno et al. method [8]	View 2	95.87	95.94	96.30	88.24	95.83	93.99
Xu et al. method [28]	View 1	97.00	98.00	98.00	94.00	99.00	99.00

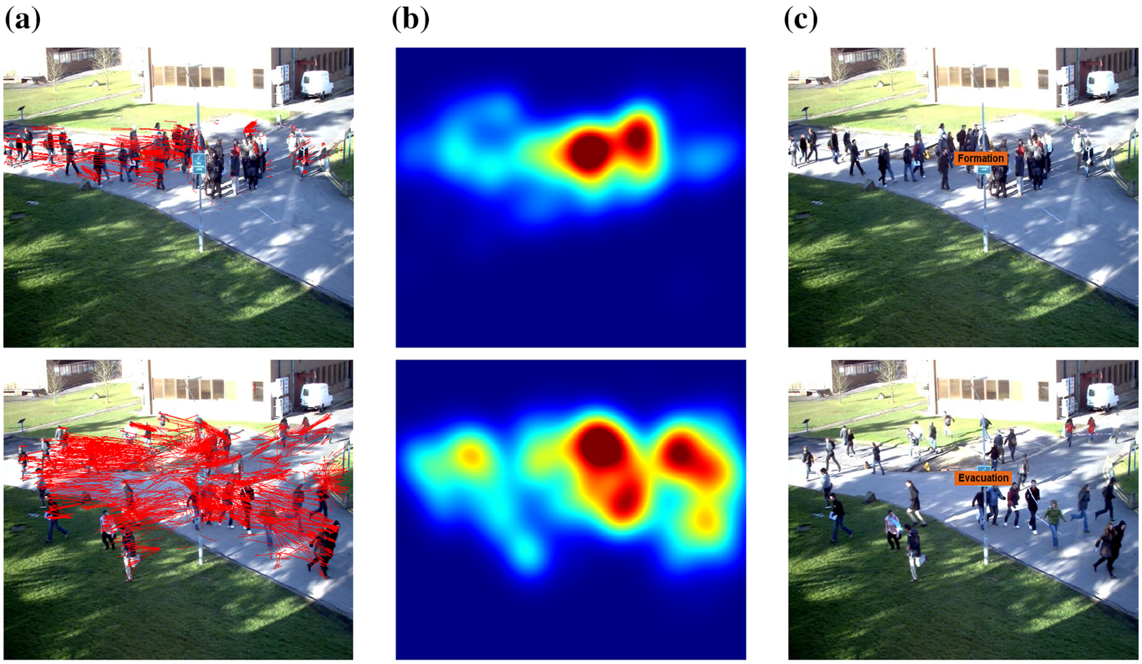

Fig. 3 Results of event characterization from PETS dataset. **a** Motion vectors, **b** density map, **c** recognized event

extracted from the crowd attributes, we are able to monitor the variation of crowd attributes in time, to interpret what is happening in the scene, to localize the event, and to have a clear idea about the density of people participating to each event. Figure 3 illustrates some examples of event characterization.

In the first row of this figure, we show a sample frame of crowd formation. This event is characterized by people coming from different directions and they are moving towards the same location (as it is depicted in the first column, showing the direction of motion vectors). Also, this event is characterized by a decrease of motion area ratio in time, in this frame it is equal to 40.72 %. In the second column, we show the estimated density map, which localizes where the crowd is formed. The area of dense regions is augmenting in time, it reaches 6.10 % at this frame. Given all the characteristics, crowd formation event is recognized and localized as it is shown in the third column.

In the second row, we show an example of evacuation. This event is characterized by the divergence of motion vectors as it is shown in the first column, because people are moving away from each others in different directions. In addition this event is characterized by a sudden increase in the speed; the average of magnitude of all motion vectors at this frame is equal to 12.05 pixels (the effects of perspective distortions are considered in the computation). This event is also characterized by in an increase in the motion area ratio (54.66 %) and a decrease in time of dense areas (as it is shown in the second column).

8 Conclusion

In this paper, we proposed a novel approach to automatically detect abnormal crowd change and to recognize crowd events in video sequences based on analyzing some attributes of feature tracks. In addition to the increasing need for automatic detection and characterization of crowd events, our study is motivated by the necessity to imply density estimation in the process because the risk of dangerous events increases when a large number of persons is involved. The effectiveness of using local density together with motion information has been experimentally validated using videos from different crowd datasets. The results show good performance for early detection of crowd change, accurate event recognition and better video understanding.

There are several extensions of this work: First, because crowd events have temporal structure, Hidden Markov Models (HMM) can tackle this classification better than SVM (classification per-frame which disregards temporal order) by capturing temporal patterns in the data. The small size of PETS.S3 dataset impeded us to investigate more this method, since HMM requires extensive training data. Another future direction of this work could be the use of the same input (local features tracking) to study group behaviors by applying trajectory clustering. Also, for change detection, our proposed method succeeds to achieve accurate results for early detection once the change occurs, however, it important to investigate event prediction before it happens.

References

1. University of minnesota crowd activity dataset. http://www.mha.cs.umn.edu/Movies/Crowd-Activity-All.avi
2. Albiol A, Silla MJ, Albiol A, Mossi JM (2009) Video analysis using corner motion statistics. In: IEEE International Workshop on PETS, pp 31–37
3. Albiol A, Silla MJ, Albiol A, Mossi JM (2009) Video analysis using corners motion analysis. In: International Workshop on Performance Evaluation of Tracking and Surveillance, pp 31–38
4. Ali S, Shah M (2007) A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, 7th edn. In: CVPR, pp 1–6
5. Almeida IR, Jung CR (2013) Change detection in human crowds. In: Conference on Graphics, Patterns and Images, p 26
6. Briassouli A, Kompatsiaris I (2011) Spatiotemporally localized new event detection in crowds. In: Proceedings of the IEEE International Conference on ICCV Workshops, pp 928–933
7. Butenuth M, Burkert F, Schmidt F, Hinz S, Hartmann D, Kneidl A, Borrmann A, Sirmacek B (2011) Integrating pedestrian simulation, tracking and event detection for crowd analysis. In: ICCV Workshops, pp 150–157
8. Cermeno E, Mallor S, Siguenza J (2013) Learning crowd behavior for event recognition. In: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), pp 1–5
9. Chan A, Morrow M, Vasconcelos N (2009) Analysis of crowded scenes using holistic properties. In: Proceedings of the 11th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS-09
10. Chan AB, Morrow M, Vasconcelos N (2009) Analysis of crowded scenes using holistic properties. In: IEEE International Workshop on PETS
11. Chen DY, Huang PC (2011) Motion-based unusual event detection in human crowds. J Visual Commun Image Represent 22(2):178–186
12. Dee H, Hogg D (2004) Detecting inexplicable behaviour. In: Proceedings of the British Machine Vision Conference, The British Machine Vision Association, pp 477–486
13. Ferryman J, Shahrokni A (2009) Pets 2009: dataset and challenge. In: PETS, pp 1–6
14. Fradi H, Dugelay JL (2012) Low level crowd analysis using framewise normalized feature for people counting. In: IEEE International Workshop on Information Forensics and Security
15. Fradi H, Dugelay J-L, (2013) Crowd density map estimation based on feature tracks. In: MMSP 15th International Workshop on Multimedia Signal Processing, September 30–October 2, 2013, Pula, Italy
16. Fradi H, Zhao X, Dugelay JL (2013) Crowd density analysis using subspace learning on local binary pattern. In: ICME 2013, IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety
17. Garate C, Bilinski P, Bremond F (2009) Crowd event recognition using HOG tracker. In: Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter), Snowbird, UT, États-Unis, pp 1–6
18. Ihaddadene N, Djeraba C (2008) Real-time crowd motion analysis. In: IEEE ICPR, pp 1–4
19. Kaltsa V, Briassouli A, Kompatsiaris I, Strintzis MG (2012) Timely, robust crowd event characterization. In: ICIP, pp 2697–2700
20. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: CVPR, pp 935–942
21. Polus A, Schofer JL, Ushpiz A (1983) Pedestrian flow and level of service. J Transp Eng 109:46–56
22. Rosten E, Porter R, Drummond T (2010) Faster and better: a machine learning approach to corner detection. IEEE Trans Pattern Anal Mach Intell 32:105–119
23. Saxena S, Brémond F, Thonnat M, Ma R (2008) Crowd behavior recognition for video surveillance. In: Proceedings of the 10th International Conference on Advanced Concepts for Intelligent

Vision Systems, ACIVS '08. Springer, Berlin, Heidelberg, pp 970–981

24. Senst T, Eiselein V, Evangelio RH, Sikora T (2011) Robust modified l2 local optical flow estimation and feature tracking. In: IEEE Workshop on Motion and Video Computing (WMVC), pp 685–690
25. Senst T, Eiselein V, Sikora T (2012) Robust local optical flow for feature tracking. Trans Circuits Syst Video Technol 22(9):1377–1387
26. Shi Y, Gao Y, Wang R (2010) Real-time abnormal event detection in complicated scenes. In: Proceedings of the 2010 20th International Conference on Pattern Recognition, ICPR '10. DC, USA, IEEE Computer Society, Washington, pp 3653–3656
27. Tomasi C, Kanade T (1991) Detection and tracking of point features. Technical report CMU-CS-91-132, CMU
28. Xu T, Peng P, Fang X, Su C, Wang Y, Tian Y, Zeng W, Huang T. (2012) Single and multiple view detection, tracking and video analysis in crowded environments. In: AVSS. IEEE Computer Society, pp 494–499
29. Zhang Y, Qiny L, Yao H, Xu P, Huang Q (2013) Beyond particle flow: bag of trajectory graphs for dense crowd event recognition. In: ICIP