Sparse Feature Tracking for Crowd Change Detection and Event Recognition

Hajer Fradi
EURECOM, Sophia Antipolis
France
Email: fradi@eurecom.fr

Jean-Luc Dugelay
EURECOM, Sophia Antipolis
France
Email: dugelay@eurecom.fr

***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 and event recognition 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 also 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 behavior. The experimental results on two different crowd datasets demonstrate the effectiveness of our proposed approach for early detection of crowd change and accurate results for event recognition.**

I. 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. Many accidents emphasize 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 insure assistance and emergency contingency plan.

Crowd behavior analysis covers different subproblems such as crowd change or anomaly detection [1], [2], [3], [4], and crowd event recognition [5], [6], [7], 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 crowd behavior. This category includes the Social Force Model [3] which is based on local characteristics of pedestrian motions and interactions, or trajectory-based methods [8]. 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 in analysis [1]. For this purpose, scene modeling techniques are used to capture the main features of the crowd behavior. These methods are based on extracting the dynamics of the entire scene and 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 study the crowd at a microscopic and a macroscopic levels. They inherit both properties to handle the limitations of each category of methods and to complement each others for better performance [5].

Our proposed approach is of hybrid nature since it incorporates optical flow information into the extracted local features and it examines long-term trajectories to capture both global and local attributes. The idea mainly consists of using low-level local features to represent individuals in the scene. By doing so, we avert typical problems encountered in detection and tracking of persons in high density crowds, such as dynamic occlusions and extensive clutter. Also, a feature tracking step is involved in the process to alleviate the effects of components irrelevant to the crowd. Better than, in our proposed approach, we extend motion information to form long-term trajectories which are less affected by noise.

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 forms, the used crowd density measure could be the number of persons [9], [10] or the level of the crowd [11]. 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 [12].

We consider that local density is an important cue for early detection of crowd event and it could complement crowd dynamics 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 large number of individuals has to be involved and to participate to crowd formation. Another example that could justify 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 space to move, as a result the speed slows down. These examples illustrate the need to use density as additional cue for crowd event characterization, also it helps to localize crowded regions.

To achieve an improved overall performance, the 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 for 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.

The remainder of the paper is organized as follows: Section II presents our proposed 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 Section III. In Section IV, we explain how to use these attributes in order to detect crowd change and to recognize crowd events. A detailed evaluation of our work follows in Section V. Finally, we briefly conclude and give an outlook of possible future works.

II. CROWD TRACKING

Although there are different approaches to the tracking problem, their application in videos of high dense crowds remains a challenge. Actually, crowded scenes exhibit some particular characteristics rendering the problem of multi-target tracking more difficult than in scenes with few people, for instance, the small size of a target in crowds, occlusions caused by inter-object interactions, constant interactions among individuals, and 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 the target objects in crowded videos, that justifies why conventional human detection or tracking paradigms fail in such cases. To overcome this problem, alternative solutions which consist of tracking particles [3], [6], [7] or local features [4] instead of pedestrians have been proposed. Other methods operate foreground masks and consider them as the regions of interest [2], [1], called activity area in [1].

In the following, our proposed approach for crowd tracking is presented. First, local features are extracted to infer the contents of each frame under analysis. Then, we perform local features tracking using the Robust Local Optical Flow algorithm from [13] and a point rejection step using forward-backward projection. The remainder of this section describes each of these system components.

A. Extraction of local features

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 description of the crowd. For local features, we assess Features from Accelerated Segment Test (FAST) [14]. FAST has the advantage of being able to find small regions which are outstandingly different from their surrounding pixels. In addition, it was used in [15] to detect dense crowds from aerial images and the derived results demonstrate a reliable detection of crowded regions. This feature is compared to the classic detector Good Features to Track (GFT) [16], which is based on the detection of corners containing high frequency information in two dimensions and typically persist in an image despite object variations.

B. Local features tracking

Local features tracking is performed by assigning motion information to the detected features. In our framework, we apply the Robust Local Optical Flow (RLOF) [13] [17], 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 [18], 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 of n_k trajectories in every time step k:

$$\mathcal{T}_k = \{T_1^k, ..., T_{n_k}^k| = \\ T_i^k = \{X_i(k-\Delta t_i^k), Y_i(k-\Delta t_i^k), ..., X_i(k), Y_i(k)\}\} \quad (1)$$

where Δt_i^k denotes temporal interval between the start and the current frames of a trajectory T_i^k. $(X_i(k-\Delta t_i^k), Y_i(k-\Delta t_i^k))$, and $(X_i(k), Y_i(k))$ are the coordinates of the feature point in its start and current frames, respectively.

III. CROWD EVENT ATTRIBUTES

We consider simultaneously local density, speed and orientation. These attributes are extracted within our proposed sparse feature tracking framework described in Section II. For local density, a probability density function (pdf) is computed on the positions of moving local features using a Gaussian kernel density, whereas, speed and orientation are extracted from motion vectors. An illustration of the modules of crowd attributes extraction is shown in Figure 1.

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

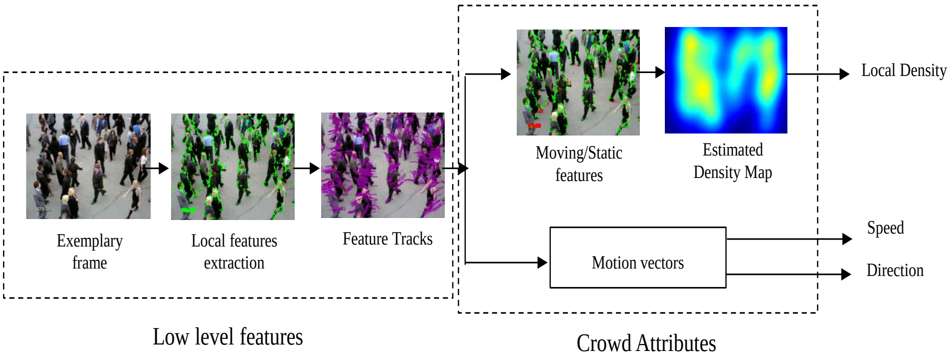


Fig. 1. Illustration of the proposed crowd attributes: crowd tracking using local features, estimation of crowd density map after distinction between moving (green) and static (red) features, estimation of speed and flow direction from motion vectors.

A. Local crowd density

Our proposed local crowd density is estimated by measuring how close local features are. This is based on the observation that the more local features come towards each other, the higher crowd density is perceived. Since the extracted local features defined in II-A 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 mean motion Γ_i^k of each trajectory T_i^k. Γ_i^k which denotes the mean of displacement between $(k - \Delta t_i^k)^{th}$ and the current frame k, is compared to a small constant ζ. Moving features are then identified by the relation $\Gamma_i^k > \zeta$ while the others are considered as part of the static background.

After filtering out static features, the crowd density map is defined as a kernel density estimate based on the positions of the moving local features. For a given video sequence of N frames $\{I_1, I_2, ..., I_N\}$, 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 as follows:

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

where σ is the bandwidth of the 2D Gaussian kernel.

The resulting crowd density map characterizes the spatial variations 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.

B. Crowd motion: Speed and Orientation

The feature tracks defined in II 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, the same feature tracks 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, the overall mean motion Γ_i^k is compared to a certain threshold β set empirically according to image resolution and camera perspective. 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. The advantage of using trajectories instead of computing motion vectors between two consecutive frames is that outliers are filtered out and the overall motion information is less affected by noise.

Once the set of useful trajectories is determined, we restrict the history of each 2D trajectory over last few frames because otherwise by considering the whole trajectory an augmentation in the speed will not be detected early and also the flow direction might be less precise. For speed estimation, it is computed 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 the trajectory.

IV. Abnormal change detection and event recognition

Overall, the spatio-temporal crowd measures introduced by density maps and motion vectors convey rich information about the distributions and the movements of pedestrians in the scene which are strongly related to their behaviors. To perform that, we model the crowd attributes by histograms, see paragraph IV-A. Then, the application of these attributes for crowd behavior analysis is demonstrated in two steps: First, the variations of a measure of stability in time is used to detect change or abnormal event, see paragraph IV-B. Second, a feature vector concatenating these histograms is employed for event recognition, see paragraph IV-C.

A. Crowd modeling

Each crowd attribute is encoded by a 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 [19] 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 [2], the speed is quantized into $N_s = 5$ classes: very slow, waking, walking fast, running, and running fast. It is important to note that speed changes can be also affected by perspective distortions, due to the fact that when people are getting away from the camera, their motion vectors are of small lengths. That is why, we rectify these effects on the speed.

B. 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 [2]: we compare the density and the motion patterns at each frame with 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:

$$S_i(k) = (C(H_i(k), H_i(k - \Delta t_1)), \\ C(H_i(k), H_i(k - \Delta t_2)), ..., C(H_i(k), H_i(k - \Delta t_n))) \quad (3)$$

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

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

where $\overline{H}$ is the mean value of H.
Similar to [2], we define the temporal stability $\sigma_i(k)$ of each histogram $H_i(k)$ as the weighted average of $S_i(k)$:

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

λ denotes the decay constant, $\Delta t_j = j\Delta t$ (Δt is a constant).

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, and 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 $(k - \Delta t_n)$:

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

C. Event Recognition

The proposed crowd attributes are also used to recognize crowd events. In particular, 6 crowd events are modeled; walking, running, evacuation, local dispersion, crowd formation and crowd splitting. In our framework, 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:

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

where θ^* are learned from the training data. This can be performed by SVM classification, and for the feature vector, we concatenate the 3 histograms $H_d(k)$, $H_\Theta(k)$, and $H_s(k)$ into $\mathcal{H}_k$. For classification, we use Chi-Square kernel:

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

V. Experimental Results

A. Datasets

First, for crowd change detection, we test our proposed approach on the publicly available UMN dataset [20], which has been widely used to distinguish between normal and abnormal crowd activities. The 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 events such as running in one direction, or people dispersing from a central point.

For the ground truth, as noticed in [2], [4], the labels of abnormal events shown in the videos are not accurate. There are some lags in the ground truth labels. To overcome this conflict, we use the labels of change detection of some videos from [2], [4], for the other videos we follow the same annotation strategy; we manually label the frame in which the crowd change occurs (for UMN dataset, once people start running).

For evaluating crowd event recognition, we test our method on PETS dataset [21], mainly on section S3, used to assess event recognition algorithms. This dataset comprises 4 video sequences of time-stamps 14:16, 14:27, 14:31 and 14:33, only the first view is used in our experiments. As noticed in [5], some sequences are composed of 2 video clips, this is the case of 14:16, 14:27, and 14:33, which results 7 videos in total. These videos depict 6 classes of crowd events: walking, running, formation (merging), splitting, evacuation, and dispersion. We annotate these videos with the 6 classes as it is shown in the following Table I.

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 I. Ground truth for event reconigition: The time intervals indicate where a specific event is recognized (from its first frame to the last one)

B. Experiments and Analysis

For quantitive evaluation of crowd change detection, we employ the relative mean frame error metric proposed in [6]:

$$e_F = N_e / N_{fr} \quad (9)$$

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

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

TABLE II. COMPARISON OF OUR DETECTION RESULTS WITH THE GROUND TRUTH LABELS USING ERROR FRAME METRIC

As demonstrated in Table II, the comparison of our detection results with the ground truth labels shows satisfactory performance and rather accurate in most videos. In terms of e_F metric, the error is small in most videos. In some cases, the delay of some frames after that the event occurs because of our strategy of detection, in which an abnormal event is detected if the temporal stability is below the dynamic threshold (defined as half the average of temporal stabilities). This requires sometimes to be detected, which justifies the delay, however, our strategy is suitable to lower false alarms.

We also compare our results with other methods, namely, the Social Force Model (SFM) [3], the adjacency-matrix based clustering (AMC) [4], and the similarity metric based on 2D-histograms decoupling speed and orientation in [2]. Figures 2, and 3 illustrate these comparisons on some videos of UMN dataset. In these figures, the green bar indicates normal events, and the red color denotes the detected or labeled abnormal event. These comparisons show that our method gives better results than SFM and comparable results regarding the two other methods. It is important to mention that UMN does not include events such as crowd formation/splitting, that could justifies that using only motion information (speed and orientation) could achieve satisfactory results. More tests on crowd events are required to demonstrate the usefulness of local crowd density as additional attribute.

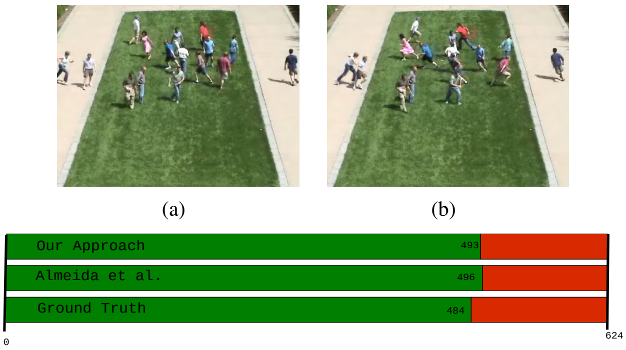
Furthermore, precision and recall of our proposed approach are listed in Table III. We compare our results to (AMC) method [4], which also runs on the same dataset and labeled the ground truth manually. The conflict concerning the ground truth annotations impeded additional comparisons. This comparison shows that our method achieves comparable results 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 [4]. The lower recall rate (of small margin) of our method, is for the same reason mentioned before about time lags in the detection until the similarity metric becomes less than the dynamic threshold.

Approach	Recall (%)	Precision (%)
Proposed approach	92.45	100
AMC approach	94	n/a

TABLE III. PERFORMANCE OF OUR PROPOSED CROWD CHANGE DETECTION METHOD IN TERMS OF RECALL AND PRECISION USING UMN DATASET COMPARED TO [4]

For crowd event recognition, we randomly split the dataset PETS. S3 into (75%) for training and (25%) for testing. This random split is done 10 times, and the following results are the average of these 10 iterations. 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 (99.54%) as classification accuracy. We also evaluate the recognition performance with confusion matrix, see Table IV.

Furthermore, we report the classification accuracy on the test set for each class separately, following one-vs-rest strategy, see Table V. As it is illustrated in these tables IV, and V, we achieve excellent results for all crowd events including crowd formation/splitting, which justifies the relevance of our proposed attributes.



(a) (b)

(c)

Fig. 2. Results on Video1 of UMN [20] dataset (a))The frame in which the crowd change occurs (b) The frame in which our method detects the crowd change (c) Comparisons of our result with [2] result and to the ground truth

VI. 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 crowd tracks. 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, and accurate event recognition.

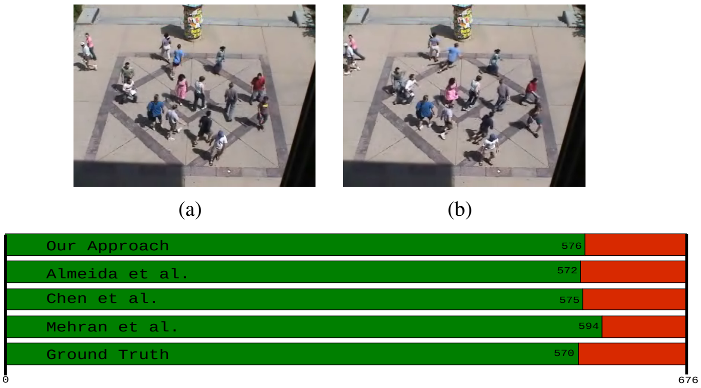
There are several possible extensions of this work: First, because crowd events have temporal structure, Hidden Markov Models (HMM) can tackle this classification better than SVM (classification per-frame which disregards temporal order) by capturing temporal patterns in the data. The small size of

	walking	running	splitting	dispersion	evacuation	formation
walking	0.9958	0.0042	0	0	0	0
running	0.0032	0.9968	0	0	0	0
splitting	0	0	1.0000	0	0	0
dispersion	0	0	0	1.0000	0	0
evacuation	0	0	0	0	0.9794	0.0206
formation	0	0	0	0.0067	0	0.9933

TABLE IV. CONFUSION MATRIX FOR EVENT RECOGNITION ON PETS2009. S3 DATASET

Events	Walking	Running	Splitting	Dispersion	Evacuation	Formation
accuracy	99.41	99.21	100.00	99.87	99.80	99.54

TABLE V. CLASSIFICATION ACCURACY OF OUR PROPOSED CROWD EVENT RECOGNITION METHOD ON TEST SET FROM PETS. S3 DATASET FOLLOWING ONE-VS-REST STRATEGY



(a) (b)

(c)

Fig. 3. Results on Video10 of UMN [20] dataset (a) The frame in which the crowd change occurs (b) The frame in which our method detects the crowd change (c) Comparisons of our result with [2], [4], [3], results and to the ground truth

PETS 2009.S3 dataset impeded us to investigate 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 performing trajectory clustering.

ACKNOWLEDGMENT

This work has received funding under the VideoSense project which is co-funded by the European Commission from the 7th Framework Programme Grant Agreement Number 261743.

REFERENCES

[1] A. Briassouli and I. Kompatsiaris, “Spatiotemporally localized new event detection in crowds.” in *ICCV Workshops*. IEEE, 2011, pp. 928–933.

[2] I. R. Almeida and C. R. Jung, “Change detection in human crowds,” in *Conference on Graphics, Patterns and Images*, no. 26, 2013.

[3] R. Mehran, A. Oyama, and M. Shah, “Abnormal crowd behavior detection using social force model,” in *CVPR*, 2009, pp. 935–942.

[4] D.-Y. Chen and P.-C. Huang, “Motion-based unusual event detection in human crowds,” *J. Visual Communication and Image Representation*, vol. 22, no. 2, pp. 178–186, 2011.

[5] C. Garate, P. Bilinski, and F. Bremond, “Crowd Event Recognition using HOG Tracker,” in *Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*. IEEE, Dec. 2009, pp. 1–6.

[6] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis, “Timely, robust crowd event characterization,” in *ICIP*, 2012, pp. 2697–2700.

[7] Y. Zhang, L. Qiny, H. Yao, P. Xu, and Q. Huang, “Beyond particle flow: Bag of trajectory graphs for dense crowd event recognition,” in *ICIP*, 2013.

[8] H. Dee and D. Hogg, “Detecting inexplicable behaviour,” in *In: Proceedings of the British Machine Vision Conference, The British Machine Vision Association*, 2004, pp. 477–486.

[9] H. Fradi and J.-L. Dugelay, “People counting system in crowded scenes based on feature regression,” in *EUSIPCO 2012, European Signal Processing Conference, August, 27-31, 2012*, 08 2012.

[10] H. Fradi and J. L. Dugelay, “Low level crowd analysis using frame-wise normalized feature for people counting,” in *IEEE International Workshop on Information Forensics and Security*, December 2012.

[11] H. Fradi, X. Zhao, and J. L. Dugelay, “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*, July 2013.

[12] H. Fradi and J.-L. Dugelay, “Crowd density map estimation based on feature tracks,” in *MMSP 2013, 15th International Workshop on Multimedia Signal Processing, September 30-October 2, 2013*, 09 2013.

[13] T. Senst, V. Eiselein, and T. Sikora, “Robust local optical flow for feature tracking,” *Transactions on Circuits and Systems for Video Technology*, vol. 09, no. 99, 2012.

[14] E. Rosten, R. Porter, and T. Drummond, “Faster and better: A machine learning approach to corner detection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, pp. 105–119, 2010.

[15] M. Butenuth, F. Burkert, F. Schmidt, S. Hinz, D. Hartmann, A. Kneidl, A. Borrmann, and B. Sirmacek, “Integrating pedestrian simulation, tracking and event detection for crowd analysis,” *ICCV Workshops*, pp. 150–157, 2011.

[16] J. Shi and C. Tomasi., “Good features to track,” in *CVPR*, 1994, pp. 593 –600.

[17] T. Senst, V. Eiselein, R. H. Evangelio, and T. Sikora, “Robust modified l2 local optical flow estimation and feature tracking,” in *IEEE Workshop on Motion and Video Computing (WMVC)*, Jan. 2011, pp. 685–690.

[18] C. Tomasi and T. Kanade, “Detection and tracking of point features,” CMU, Technical Report CMU-CS-91-132, 1991.

[19] A. Polus, J. L. Schofer, and A. Ushpiz, “Pedestrian flow and level of service,” *Journal of Transportation. Engineering*, vol. 109, pp. 46–56, 1983.

[20] “U. of minnesota crowd activity dataset,” http://www.mha.cs.umn.edu/Movies/Crowd-Activity-All.avi.

[21] J. Ferryman and A. Shahrokni, “Pets2009: Dataset and challenge,” in *PETS*, 2009, pp. 1–6.