TIMELY, ROBUST CROWD EVENT CHARACTERIZATION

Vagia Kaltsa[1,2]*, Alexia Briassouli* [2]*, Ioannis Kompatsiaris*[2]*, Michael G. Strintzis*[1]

[(1)]Aristotle University of Thessaloniki,[(2)]Informatics and Telematics Institute, CERTH

ABSTRACT

The automated analysis of crowd behavior from videos has been a rather challenging problem to address due to the complexity and density of the motion, occlusions and local noise. A novel approach for the fast and reliable detection and characterization of abnormal events in crowd motions is proposed, based on particle advection and accurate optical flow estimation. Experiments on benchmark datasets show that changes are detected reliably and faster than existing methods. Also, regions of change are localized spatially, and the events occurring in the video are characterized with accuracy.

Index Terms— crowd, event detection, particle advection

1. INTRODUCTION

The analysis of crowd behaviors finds applications in many areas, such as surveillance, security, public safety, public space design, crowd management, and the development of virtual and intelligent environments. Crowded scenes contain many occlusions, high density and complex motion, which render object detection and tracking challenging. A new spate of approaches is being developed for the effective analysis of crowds. The PETS 2009 data (http://www.cvg.rdg.ac.uk/PETS2009/a.html) and videos (http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi) from the University of Minnessota (UMN) serve for benchmarking.

One category of crowd analysis methods considers that the crowd consists of individuals to be extracted and/or tracked. These include the Social Force model [1], [2], which is based on local characteristics of pedestrian motions and interactions, or trajectory-based methods [3], [4], [5]. Such methods encounter difficulties because of the density of the crowd, dynamic occlusions and errors in tracking. Crowds are treated as a whole in "holistic" methods, which capture the dynamics of the entire scene [6], [7], [8]. Hybrid methods examine the data both at a microscopic and a macroscopic level to overcome some of the limitations of each approach [9], [10], [11].The proposed method is of a hybrid nature, as it incorporates local optical flow information via particle advection (Sec. 2.1) and examines particle clusters globally, analyzing their velocities to detect abnormal events and characterize the crowd motions. The proposed method detects abnormal events faster than existing approaches, while additionally characterizing the crowd motions and spatially localizing the regions where new events occur.

This paper is organized as follows: Section 2 describes particle advection and how clusters are formed on the warped particle grid so as to reflect the changes caused by crowd motions. Spatial localization of crowd motions, the detection of abnormal events based on cluster velocities and the characterization of the events taking place are presented in Section 3. Section 4 presents experimental results with the PETS2009 and UMN datasets, and conclusions are drawn in Section 5.

2. PARTICLE ADVECTION AND CLUSTERING

The dense optical flow information is used within a particle advection framework so as to avoid inaccuracies from occlusions or other sources of local noise common in crowded scenes. The displaced particles form clusters whose mean velocity is examined to detect and characterize new events.

2.1. Particle Advection

The crowds are considered to consist of particles that are displaced, i.e. advected, by a field consisting of the optical flow estimates. An initial symmetrical grid of "sources" is set up over each frame at $t = 0$, and each source produces a particle, creating the initial symmetrical particle grid. Then, each particle's position $\bar{p}$ is advected according to the optical flow $\bar{v}(t, \bar{p})$ at that location and at frame t:

$$\dot{\bar{p}} = \bar{v}(t, \bar{p}) \Leftrightarrow \frac{d\bar{p}}{dt} = \bar{v}(t, \bar{p}). \quad (1)$$

The particle trajectory can be found by solving the following Initial Value Problem (IVP) to find its new positions:

$$\dot{\bar{p}} = \bar{v}(t, \bar{p}), t \geq 0, \quad \bar{p}(0) = \bar{p}_0. \quad (2)$$

Eq. (2) can be solved with an Ordinary Differential Equation (ODE) solver to interpolate the flow field. Euler integration is sufficient, as higher order Runge-Kutta solvers do not give significantly different results [12]. Thus, eq. (2) is approximated by the first order Taylor expansion:

$$\bar{p}(t_0 + h) \simeq \bar{p}(t_0) + h \frac{d\bar{p}}{dt}\Big|_{t=t_0}. \quad (3)$$

A particle's position after $h = 1$ frame is given by the approximation of eq. (3) after rounding with the flooring operation:

$$\hat{p}(t_0 + 1) = \left\lfloor \bar{p}(t_0) + \frac{d\bar{p}}{dt}\Big|_{t=t_0} \right\rfloor = \lfloor \bar{p}(t_0) + \bar{v}(t, \bar{p})|_{t=t_0} \rfloor. \quad (4)$$

Thus, the initial symmetric particle grid generated by the sources is warped at each frame according to the optical flow, without being affected by occlusions or local noise. Examples of particle advection in Fig. 1 show that the particle displacements reflect the true motions in the scene. Each particle at location $\bar{p}$ is then characterized by the feature vector:

$$\bar{f}(\bar{p}) = [\bar{s}, \bar{p}, t_L, |\bar{v}|, \angle\bar{v}, n(\bar{p})], \quad (5)$$

where $\bar{s}$ is its parent source (the location from which it originates), $\bar{p}$ its current location, t_L its lifespan, $|\bar{v}|, \angle v$ are the magnitude and angle respectively of its optical flow, and $n(\bar{p})$ is a cluster id which will be necessary for clustering. Dense optical flow is estimated using the method of Farneback [13], which provides good accuracy without compromising speed, and the flow estimates are smoothed using a 3×3 Gaussian kernel to reduce noise. The particles generated by each source and their feature vectors are recorded in a list which is accessed to detect and characterize crowd events.

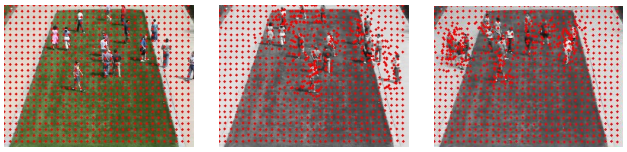

Fig. 1. Particle grid initially and after particle advection.

2.2. Particle Generation

The following steps take place for particle generation:
(1) Each particle location is updated according to eq. (4): if is out of image bounds, it dies and if it is the last particle of a source, a new one is generated.
(2) A particle's lifespan t_L depends on the desired execution time: if a particle has lived too long, slowing down the program, and is not the last particle, it dies.
(3) If a particle stops moving is not the last particle of a source, it dies. This eliminates false clusters from particles that are now immobile.
(4) Each source calculates its distance from the last particle it generated at location $\bar{p}$ via eq. (4): if it is higher than the particle distance on the initial grid, a new particle is generated, so that there is never a big gap between particles. Thus, the particle grid is updated via particle generation and advection and its features are stored in $\bar{f}(\bar{p})$ of eq. (5). When there are small motions, the particles undergo small changes and small clusters with common velocity characteristics are formed, as in Fig. 2(a) - (c), while large optical flow induced particle displacements affect greater areas and form large clusters (Fig. 2(d) - (f)).

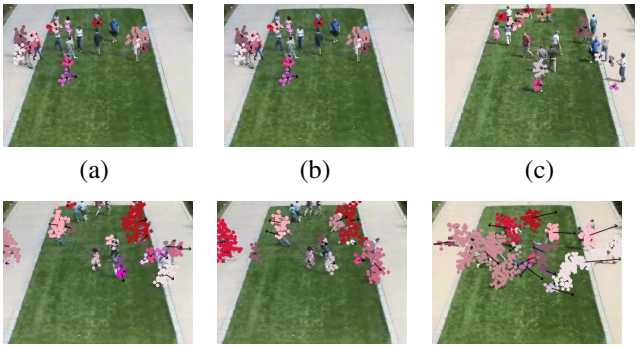

Fig. 2. (a) - (c) Clustering of particles after advection. (d) - (f) Particle clusters with the corresponding mean optical flow.

2.3. Particle Clustering

In order to form clusters from the particles that are being displaced, the density based DBSCAN algorithm is used [14], as it is robust to noise, it does not require prior determination of the number of clusters (unlike K-means), and it is capable of handling clusters of different sizes and shapes. The clustering of the particles $n(\bar{p})$ is based on their location $\bar{p}$. Each cluster contains the particle id $n(\bar{p})$ and their features $\bar{f}(\bar{p})$ of eq. (5). The average velocity $(|\bar{v}|, \angle\bar{v})$ of the particles in a cluster are used for the detection and characterization of events, which are simultaneously spatially localized in the clusters.

3. ABNORMAL EVENT LOCALIZATION AND RECOGNITION

The events examined in this work correspond to the events in the UMN and PETS2009 videos. Thus, we consider that abnormal events occur when there is a change from a crowd walking to running, and the events to be detected are run left, run right, run up, run down, merge, evacuate. Cluster direction is estimated as a mean of circular quantities instead of arithmetic mean [15], as the former is more meaningful in non-Euclidean spaces, like that of the cluster direction. The mean of circular quantities for n angles $\alpha_j, j = 1, ..., n$ is calculated by first converting all points into Cartesian coordinates. The mean of the Cartesian coordinates is estimated and converted into polar coordinates, giving:

$$\alpha = \arctan\left(\frac{1}{n}\sum_{j=1}^{n}\sin\alpha_j, \frac{1}{n}\sum_{j=1}^{n}\cos\alpha_j\right). \quad (6)$$

An example of a mean of circular quantities is shown in Fig. 3, where the mean angle is correctly found to be α after projection on the cartesian axes.

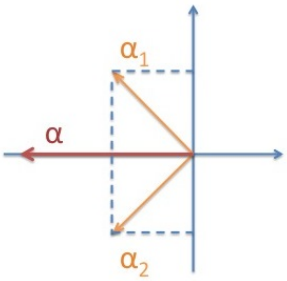

Figure 3. The mean of circular quantities gives the true mean of the angles $\alpha_1,\ \alpha_2$, while the arithmetic mean would incorrectly give zero.

3.1. Change Detection for Abnormal Event Localization

Changes are found by comparing a test quantity T_k to a video-independent threshold obtained from training data, i.e. a subsequence of the actual video formed by its first frames. In frame k with N_c clusters, the mean velocity of each cluster i ($i = 1, ..., N_c$) with N_i particles is $u_i(k) = \frac{1}{N_i} \sum_{\bar{p} \in Cluster\ i} v_i(k, \bar{p})$, where $v_i(k, \bar{p})$ is the velocity of the particle at location $\bar{p}$ in cluster i at frame k. Then:

$$T_k = \sum_{i=1}^{N_c} \sum_{\bar{p} \in Cluster\ i} v_i(k, \bar{p}). \tag{7}$$

At each frame, T_k is compared to a threshold η: if $T_k > \eta$, a change has occurred and a new event is taking place. This threshold is determined empirically vis a general formula: for N_f training frames and N_{all} particles in them , we get η:

$$\eta = \mu \cdot N_{all} + c \cdot \sigma = \sum_{k=1}^{N_f} \sum_{i=1}^{N_c} \sum_{\bar{p} \in Cluster\ i} v_i(k, \bar{p}) + c \cdot \sigma$$

where $c = 3$ has been empirically found to give the quickest and most accurate change detection results, μ is the mean of the velocities of all particles in all clusters of the training data and σ is their standard deviation:

$$\begin{aligned} \mu &= \frac{1}{N_{all}} \sum_{k=1}^{N_f} \sum_{\bar{p} \in Tr.Clust.} v_i(k, \bar{p}) \qquad (8) \\ \sigma &= \frac{1}{N_{all}} \sum_{k=1}^{N_f} \sum_{\bar{p} \in Tr.Clust.} v_i^2(k, \bar{p}) - \mu^2. \end{aligned}$$

When there are many particles in a frame, true changes in correspond to a much higher T_k, so N_{all} is included in Eq. (8).

3.2. Event Characterization

In order to characterize the events mentioned above, we first examine if there is motion in a dominant direction via voting. If there is a large cluster in a frame, small clusters are ignored, while the rest give as many votes for their mean velocity direction as their cardinality. Directions are determined by the angle α of each cluster's mean velocity: (1) $-\pi/4 \leq \alpha \leq \pi/4$: right, (2) $\pi/4 \leq \alpha \leq 3\pi/4$: up, (3) $3\pi/4 \leq \alpha \leq \pi$, $-\pi \leq \alpha \leq -3\pi/4$: left, (4) $-3\pi/4 \alpha \leq -\pi/2$ and the majority vote determines if there is a dominant motion direction. If there is no large cluster, all clusters "vote", and if three of the four directions have votes in the 30^{th} percentile of the fourth direction's votes, the latter is considered to dominate the scene. When no motion in a specific direction is found, we examine if "merge" or "evacuation" occurring. The frame is divided into four quarters, and each cluster is examined: if it is in the first quarter and is moving up or right, in the second quarter moving up or left, in the third quarter moving down or left, in the fourth quarter moving down or right, we detect evacuation. All other cases correspond to merging. This voting scheme provides very accurate motion classification. The entire crowd event recognition process is depicted in Fig. 4.

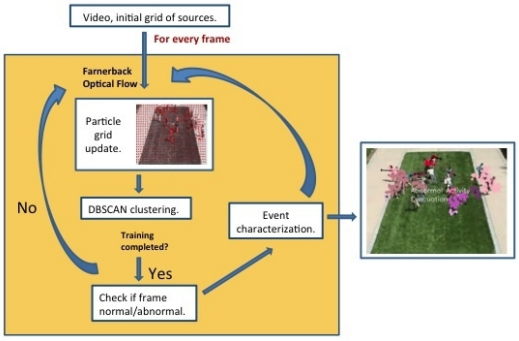

Fig. 4. Proposed method for abnormal crowd event detection

4. EXPERIMENTS

Experiments take place with the benchmark videos from UMN and PETS2009. The time when an abnormal event occurs is detected, events taking place after a change are characterized according to Sec. 3 and comparisons with ground truth show that they are correctly recognized. For a quantitative evaluation, we define the relative mean frame error for each video with N_{fr} frames and N_e error frames as:

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

Table 1 shows that the proposed method is accurate and in most cases faster than the work at UMN (http://mha.cs.umn.edu/proj_events.shtml#crowd), while frames with detected abnormal events are shown in Fig. 5. The numerical differences in the detected changes appear small, but are actually important for timely detection in critical situations. This can be better understood by seeing the videos of change detection and crowd motion characterization in http://mklab.iti.gr/content/abnormal-event-detection, where our results are compared to those of http://mha.cs.umn.edu/proj_events.shtml#crowd.

Dominant Path Localization: For each source, the ids of the particles it generated are saved. In regions with considerable motion, many particles are generated by the underlying source, so their number can be determine areas of significant activity. Fig. 6 shows examples of dominant

Seq. PETS	Ground truth	Det. Changes	e_F
14-16 view 1	42, 173	42, 184	0.025
14-31 view 1	90	90	0
14-33 view 1	342	349	0.054
Seq. UMN	Ground Truth	Ours, e_F	UMN, e_F
abnormal1a	414	418, 0.008	441, 0.044
abnormal1b	570	575, 0.007	591, 0.023
abnormal2a	275	276, 0.002	296, 0.065
abnormal2b	485	496, 0.019	503, 0.012
abnormal2c	425	427, 0.003	445, 0.028
abnormal2d	386	411, 0.039	403, 0.017
abnormal2e	629	629, 0	645, 0.021
abnormal2f	399	408, 0.016	412, 0.007
abnormal3a	460	464, 0.069	502, 0.077
abnormal3b	487	489, 0.004	525, 0.064
abnormal3c	607	610, 0.004	642, 0.048

Table 1. New event detection in UMN, PETS2009. The proposed method detects changes correctly and faster than UMN.

Fig. 5. Abnormal event detection in the PETS, UMN datasets

paths in the PETS and UMN videos, where busiest regions are colored purple, then red, green, blue. More examples can be found at http://mklab.iti.gr/content/dominant-path-localization-crowd-videos.

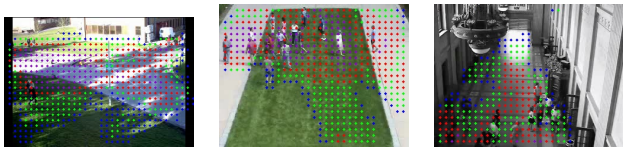

Fig. 6. Dominant paths in the PETS, UMN datasets

5. CONCLUSIONS

A novel method for the detection of new events in crowds is proposed, based on dense optical flow and particle advection, to avoid occlusion-related inaccuracies. DBSCAN-based clustering of displaced particles helps reliably determine regions with similar velocities, whose changes correspond to abnormal events. A voting scheme characterizes crowd motions by detecting if there are dominant crowd motions or if events like evacuation and merging are taking place. Comparisons with existing methods show we achieve faster, accurate event detection, with simultaneous spatial localization of the activities and the dominant path. Future work includes examining complicated videos, involving complex crowd motions.

Acknowledgement: This work was funded by the European Commission under the 7th Framework Program (FP7 2007-2013), grant agreement 288199 Dem@Care.

6. REFERENCES

[1] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical Review*, vol. 51, no. 4282, 1995.

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

[3] H. Dee and D. Hogg, "Detecting inexplicable behaviour," in *Proc. British Machine Vision Conference*, 2004.

[4] N. Ihaddadene and C. Djeraba, "Real-time crowd motion analysis," in *ICPR,19th International Conference on Pattern Recognition (ICPR 2008), December 8-11, 2008, Tampa, Florida, USA*, 2008, pp. 1–4.

[5] Jialue Fan, Nan Jiang, and Ying Wu, "Automatic video-based analysis of animal behaviors," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, sept. 2010, pp. 1513 –1516.

[6] S. B. Ernesto, Andrade, and R. B. Fisher, "Modelling crowd scenes for event detection," in *ICPR*, 2006, pp. 175–178.

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

[8] A. Briassouli and I. Kompatsiaris, "Spatiotemporally localized new event detection in crowds," in *ARTEMIS ICCV 2011, International Conference on Computer Vision*, Nov. 2011, pp. 928 – 933.

[9] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *IEEE Int. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[10] A. Albiol, M. J. Silla, A. Albiol, and J. M. Mossi, "Video analysis using corner motion statistics," in *PETS, IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 31–37.

[11] C. Garate, P. Bilinski, and F. Brmond, "Crowd event recognition using HOG tracker," in *Proceedings of the Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, (invited paper), Winter-PETS 2009*, Dec. 2009.

[12] A. Treuille, S. Cooper, and Z. Popovic, "Continuum crowds," *ACM Trans. Graph*, vol. 25, pp. 1160–1168, 2006.

[13] G. Farnebäck, "Fast and accurate motion estimation using orientation tensors and parametric motion models," in *Proceedings of 15th International Conference on Pattern Recognition*, 2000, ICPR'00, pp. 363–370.

[14] M. Kryszkiewicz and L. Skonieczny, "Faster clustering with DBSCAN," in *Intelligent Information Processing and Web Mining*, M. Klopotek, S. Wierzchon, and K. Trojanowski, Eds., vol. 31, pp. 605–614. Springer Berlin/Heidelberg, 2005.

[15] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics, 2007.