Change detection in human crowds

Igor Rodrigues de Almeida, Claudio Rosito Jung
Institute of Informatics - Federal University of Rio Grande do Sul
Email: iralmeida@inf.ufrgs.br, crjung@inf.ufrgs.br

***Abstract*—This paper presents a method to detect unusual behavior in human crowds based on histograms of velocities in world coordinates. A combination of background removal and optical flow is used to extract the global motion at each image frame, discarding small motion vectors due artifacts such as noise, non-stationary background pixels and compression issues. Using a calibrated camera, the global motion can be estimated, and it is used to build a 2D histogram containing information of speed and direction for all frames. Each frame is compared with a set of previous frames by using a histogram comparison metric, resulting in a similarity vector. This vector is then used to determine changes in the crowd behavior, also allowing a classification based on the nature of the change in time: short or long-term changes. The method was tested on publicly available datasets involving crowded scenarios.**

***Keywords*-Human crowd analysis; Unusual event detection; Video surveillance**

I. INTRODUCTION

Significant effort has been devoted to video surveillance and human motion understanding in the past years, both in the industry and the academy. Nowadays, there commercial systems developed to track (e.g. www.smarteye.se), recognize (e.g. www.iteris.com) and understand the behavior of a great variety of objects, using one or multiple video cameras, processing the information in one or more computers. In the academia, there are several recent papers on surveillance-related topics, as illustrated in [1].

In particular, several real-life situations involve crowded scenarios. Crowds may arise in busy streets, sporting events, music concerts, among others, and small disturbances may lead to a panic situation and possibly tragic consequences. When dealing with crowds, conventional computer vision methods for person identification/tracking individuals are not appropriate due to severe clutter and occlusions, as related by Zhan et al. [2]. Furthermore, studies conducted by psychologists and sociologists have pointed out that in several crowded scenes, people can lose their individualities and adopt the behavior of the crowd entity, behaving in a different way than if they were alone [3]. It means mainly that a collective entity can emerge, depending on many aspects such as people goals, the observed environment, the occurred event, as well as other variables. This type of collective analysis can be useful to measure the level of crowd comfort and to detect some specific events, as described in [4].

This paper presents a new approach for change detection in crowded scenes based on the temporal analysis of perspective-corrected optical flow vectors, which may possibly be related to the emergence of unusual behavior. The remainder of this paper is organized as follows: Section II revises the state-of-the-art on vision-based crowd analysis; Section III presents the proposed approach, and Section IV shows the experimental results; finally, the conclusions are presented in Section V.

II. RELATED WORK

There are several approaches for analyzing crowd behaviors using computer vision, which can be divided into three main classes: microscopic, macroscopic, and a combination of the two [5]. In the microscopic approach people are analyzed as discrete individuals, and this information is used to infer the behavior of the crowd. In the macroscopic approach the crowd is instead analyzed as a single unit, with no individual pedestrian detection/tracking, which is a way of avoiding the problems with occlusion. A combination of micro- and macroscopic approaches can be made by keeping the crowd as a homogeneous mass, but at the same time considering an internal force. Another way is by keeping the characters of the people while maintaining a general view of the entire crowd. Next we review some methods for crowd analysis using computer vision.

Mehran et al. [5] used social force model to detect and localize unusual behavior in crowded scenes. In their approach, the interaction of particles guided by a space-time average of the optical flow are estimated using social force model, and a bag of s approach is adopted for unusual event detection (randomly selected sets are used to model normal behavior). Solmaz and colleagues [6] also used optical flow to guide particles, but explored concepts related to the stability of dynamical system to detect pre-determined events in a crowd. Ihaddadene and Djeraba [7] also explored optical flow algorithms, but limited to regions of interest based on a motion intensity heat map. Within these regions of interest, sudden changes and abnormal motion variations are detected based on velocity, direction and density information.

Dee and Caplier [8] presented a prototype system for the automated analysis of crowded scenes based on local histograms of motion vectors. After detecting pedestrians and faces to estimate scale, they use the KLT tracker [9] to obtain pieces of trajectories, which are used to estimate local motion. Histograms of local motion are computed regionally by dividing each frame into a set of square regions, and compare the histograms of each frame to the average histograms of a training set. Li and colleagues [10] also presented a histogram-based approach for crowd analysis, but focusing on movement segmentation. In their approach, optical flow is used to detect foreground moving pixels, and a histogram of the flow vector angles is built (and smoothed). Then, modes of the smoothed histogram are used to segment the flow. Clearly, their approach is only sensitive to angular variations. The method proposed in [11] combines crowd kinetic energy, motion variation and direction variation for the abnormality detection. The motion variation is derived from the crowd kinetic energy of two adjacent frames, and the motion direction variation is estimated using mutual information of the direction histograms of two neighboring motion vector fields.

In [12], Brostow and Cipolla use an unsupervised data driven Bayesian algorithm that has detection of individual entities as its primary goal to observe the information of motion of the individuals. They track image features and group them using a probabilistic approach, where each group represent the movement of an entity.

In [13], Chen and Huang used optical flow to cluster human crowds into groups in an unsupervised manner using a novel approach, called adjacency-matrix based clustering. Each cluster is characterized based on the social force model, and unusual crowd events are detected when the orientation of a crowd is abruptly changed or when interaction among crowds is not similar to the predicted value.

Briassouli and Kompatsiaris [14] presented an approach based on properties of the data in the Fourier domain for detecting new events in crowds. Their method does not require extensive training, estimation of the optical flow or data modeling. In fact, random crowd motion is encoded in the phase of the Fourier coefficients, and statistical sequential change detection methods (e.g. the Cumulative Sum) are applied to detect events in crowds.

Haque and Murshed [15] presented a new approach for handling crowd scenarios that is based neither on motion cues nor trajectories. Instead, the explored feature extraction based on frame-set characteristics computed on foreground blobs. The temporal variation of frame-level features is analyzed over sliding temporal window, and a set of specific events is trained using Support Vector Machines (SVMs).

The method presented by Wu and colleagues in [16] aims at detecting and localizing anomalies in complex and crowded sequences by using a Lagrangian particle dynamics approach, together with chaotic modeling. Also, representative trajectories are defined to serve as a compact modeling elements in crowd flows. Representative trajectories also provide a simple way of obtaining time series data, which can effectively be used for chaotic modeling of a scene. The representative chaotic feature set is regulated to reliably capture the chaotic dynamics of representative trajectories to be used for probabilistic anomaly detection and localization.

Andersson and colleagues [17] presented an approach for detecting anomalous motion patterns in crowds. The authors use K-means clustering for identifying groups and Hidden Markov Models (HMMs) for modeling the expected motion patterns of dense and calm groups. Although the experimental results shown in the paper are good, the scenarios used in the experiments are not very dense.

As it can be observed, most existing approaches explore motion cues of the crowd. However, when camera perspective effects are significant (camera far from top-view setup), the same motion vectors in world coordinates may map to very different motion vectors in image coordinates. In [8], a rough scale estimator is performed by detecting pedestrians, which also may fail when perspective effects are strong. Also, the use of training sets or pre-defined set of events limits the practical application of these methods.

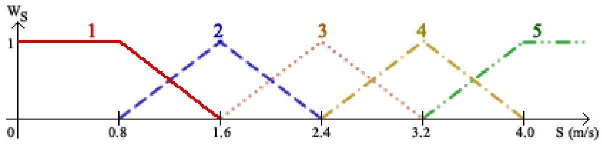
This paper presents an approach for detecting changes in the global crowd motion behavior based on motion vectors in world coordinates, do not require a set of training frames. In fact, each new frame is compared to a set of previous frames, and motion-based behavioral changes can be detected and classified as short-term (abrupt) or long-term (smooth).

III. The Proposed Method

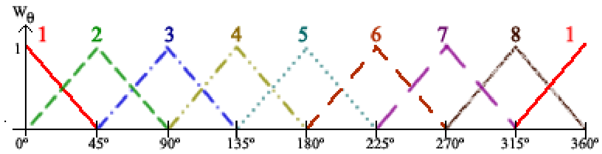
Let us consider a calibrated static surveillance camera, and assume that the filmed region is roughly planar (the ground plane is given by $z = 0$). The proposed approach starts extracting foreground blobs using the approach presented in [18]. This method builds a background model and local estimates of the noise using robust statistics, so that it may be applied even when strong motion is present in the set of frames used to learn the background model. Also, shadow removal is included in [18], reducing the number of false detections.

We also estimate the crowd movement in the scene using a robust optical flow algorithm [19]. This method employs a variational approach that can cope with large displacement vectors, common in low framerate surveillance cameras, generating a vector field $\boldsymbol{v}(\boldsymbol{x})$. Since artifacts present in the video stream (such as noise, moving shadows, waving trees, etc.) generate a series of spurious motion vectors, we restrict the output of the optical flow $\boldsymbol{v}(\boldsymbol{x})$ only in to foreground pixels.

At this stage, we have a binary mask $F(\boldsymbol{x})$ with pixels that belong to the foreground (assumed to be members of the crowd), along with the corresponding motion vectors $\boldsymbol{v}(\boldsymbol{x})$ (in pixels) obtained with the optical flow. However, two people moving with the exact same velocity in world coordinates can have different motion vectors estimated with optical flow, due to camera perspective. To estimate the velocity vectors in world coordinates, it is necessary to know the camera matrix and the height (z coordinate) of each pixel, and then compute the inverse perspective mapping to obtain the motion vector (x, y) on the ground plane (in meters) based on pixel coordinates (u, v). Since it is difficult to provide an estimate of the pixel height (but plausible to assume that it lies in the range $[0, h_{max}]$, where h_{max} is the maximum height allowed to a person, e.g. set to 1.75m), we compute the inverse mapping using the ground plane homography (i.e., we assume that $z = 0$). Although computing the inverse mapping with $z = h_{max}/2$ could reduce the approximation error, we adopted $z = 0$ since the ground plane homography can be computed easily by selecting points on the ground plane with known world coordinates, or based on the motion of pedestrians [20].



(a)



(b)

Fig. 1. Speed (a) and orientation (b) weighting functions used to obtain the 2D motion histograms.

Given the velocity vector field $\boldsymbol{v}_w(\boldsymbol{x})$ in world coordinates at a given frame t, the global motion of the crowd is encoded by a 2D histogram, decoupling speed and orientation. For each pixel $\boldsymbol{x}$ related to a person in the crowd, we quantize the speed $s(\boldsymbol{x})$ (in m/s, estimated based on the frame rate of the video sequence) and the orientation $\theta(\boldsymbol{x})$ (in degrees) into N_s and N_θ bins, respectively. We have chosen $N_s = 5$ to quantize the speed into five speed classes: very slow, walking, walking fast, running and running fast, using a bin size of $\Delta s = 1.6$ m/s, so that the transition between "walking fast" and "running" occurs at a speed of about 2.4 m/s, in agreement with [21]. If only image coordinates are used, the definition of the bin sizes related to speed becomes a complex problem, particularly when perspective effects are more noticeable. As for the quantization of the orientations, we have defined experimentally $N_\theta = 8$, based on cardinal and ordinal directions, leading to an orientation bin size $\Delta\theta = 45°$.

To reduce influence of noise and quantization issues, we first estimate the underlying probability distribution function (PDF) using Kernel Density Estimation (KDE) [22] instead of computing the histogram directly from $\boldsymbol{v}_w(\boldsymbol{x})$. In KDE, a kernel centered at each observation is used to obtain a continuous PDF of the data, spreading its influence throughout more than one histogram bin. We have used triangular kernels in both speed and orientation dimensions of the histogram. The supports of the two triangular windows were defined as the speed and orientation bin sizes (Δs and $\Delta\theta$, respectively) so that each velocity vector $\boldsymbol{v}_w$ with speed s and orientation θ contributes to the histogram H according to

$$H(i,j) = \max\left\{0, 1 - \frac{|s - s_i|}{\Delta s}\right\} \cdot \max\left\{0, 1 - \frac{|\theta - \theta_j|}{\Delta\theta}\right\}, \tag{1}$$

where s_i and θ_j are the centers of the speed and orientation bins, respectively, and i, j the bin indices. It can be observed that only the four bins closest to (s, θ) present non-zero weight. For the sake of illustration, the triangular weighting functions in s and θ are shown in Fig. 1.

The pipeline for estimating the KDE-weighted histograms is illustrated in Fig. 2. A typical frame is shown in Fig. 2(a), and the result of the background removal approach is illustrated in Fig. 2(b). The optical flow vectors computed for valid foreground pixels are shown in Fig. 2(c), and the 2D histogram is illustrated as a surface in Fig. 2(d).

A. Crowd Behavior Analysis

For each frame t, we compute the speed-orientation histogram according to the procedure described so far (in fact, to include some temporal smoothness into the histograms, we use samples from both frames t and $t-1$ to build the histogram). We then normalize the histograms (so that they can be treated as discrete PDFs), obtaining normalized histograms $H_t(i,j)$. If the motion pattern of the crowd remains similar within a time period, the corresponding histograms are expected to be similar. On the other hand, changes in the crowd behavior are expected to generated discrepancies when comparing the histograms.

In the proposed approach, instead of generating a training set to learn "usual motion", we compare the motion at each frame with the motion patterns in a set of previous frames. More precisely, we generate a similarity vector $\boldsymbol{S}_t$ given by

$$\boldsymbol{S}_t = \Big(C(H_t, H_{t-\Delta t_1}),\ C(H_t, H_{t-\Delta t_2},\ ..., C(H_t, H_{t-\Delta t_n}) \Big), \tag{2}$$

where n is the number of previous frames used in the comparison, Δt_i are the frame steps, and C is the histogram similarity metric. Although there are several possibilities for C, we have used the histogram correlation, which produces values between 0 and 1. Formally, the correlation between histograms H_1 and H_2 is given by

$$C(H_1, H_2) = \frac{\displaystyle\sum_{i,j}(H_1(i,j) - \overline{H_1}) \sum_{i,j}(H_2(i,j) - \overline{H_2})}{\sqrt{\displaystyle\sum_{i,j}(H_1(i,j) - \overline{H_1})^2} \sqrt{\displaystyle\sum_{i,j}(H_2(i,j) - \overline{H_2})^2}} \tag{3}$$

where $\overline{H}$ is the mean value of H.

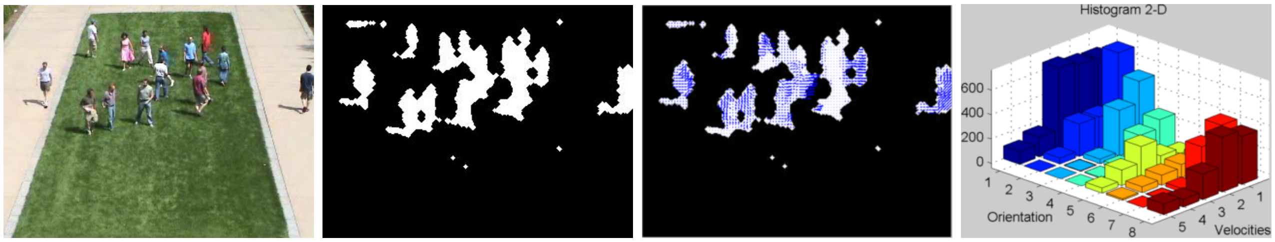
The similarity vector $\boldsymbol{S}_t$ is then analyzed to detect changes in the crowd behavior, as well as to classify it in terms of how fast it has occurred. The temporal stability σ_t of the crowd behavior at frame t is defined as a weighted average of $\boldsymbol{S}_t$:

$$\sigma_t = \boldsymbol{w}^T \boldsymbol{S}_t, \tag{4}$$

where $\boldsymbol{w}$ is the weighing vector that presents larger values for recent frames. We have used exponentially decaying weights

$$\boldsymbol{w} = \frac{1}{\displaystyle\sum_{i=1}^{n} e^{-\lambda\Delta t_i}} \left(e^{-\lambda\Delta t_1},\ e^{-\lambda\Delta t_2}, \cdots, e^{-\lambda\Delta t_n} \right), \tag{5}$$

where λ is the decay constant. In all experiments, we have used $n = 14$ frames to build $\boldsymbol{S}_t$, and used $\Delta t_i = i\Delta t$, with Δt constant and set to 0.57s, so that a time period of approximately 8s is evaluated. This value for Δt corresponds to 4 frames when the sequence is acquired at 7 frames per



(a) (b) (c) (d)

Fig. 2. (a) Selected frame. (b) Foreground pixels. (c) Optical flow vectors at foreground pixels. (d) Surface illustrating the 2D histogram $H(i,j)$.

second, which is common for surveillance cameras. The decay constant λ was set to $\lambda = 0.52$ experimentally.

A behavior change is detected when the temporal stability σ_t is low, meaning that the similarity between the current frame and the previous ones is small. More precisely, we define an adaptive threshold β_t based on the history of σ_t:

$$\beta_t = \frac{1}{2n}\sum_{i=1}^{n} \sigma_{t-\Delta t_i}, \tag{6}$$

i.e. the threshold is half the average of the temporal stability values σ_t in a temporal window.

Fig. 3 shows an example of crowd behavior change for the PETS2009 S3.Event Recognition dataset[1], sequence 4. In this sequence, people enter the scene from several directions, and gather at the center. After some time, they all start running almost instantly away from the center, each one in a randomly direction. In the bottom of Fig. 3 we show the values of σ_t along the video sequence. We can see that σ_t oscillates even while the crowd behavior does not change. When change in the crowd behavior occurs, σ_t drops sharply, and the detected unusual behavior is indicated by a vertical red line.

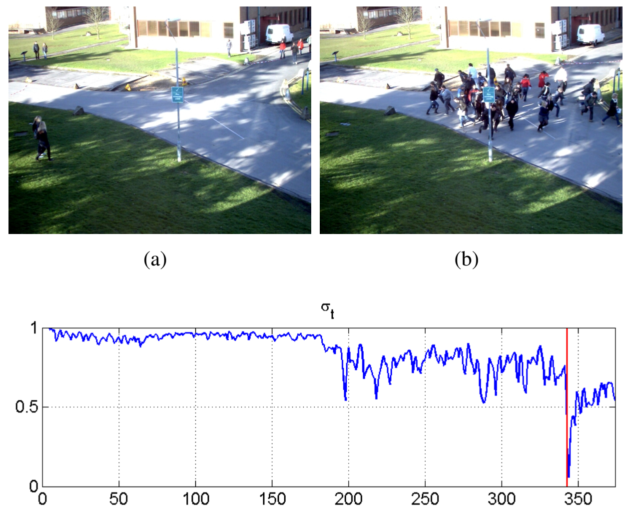
When a change in behavior is detected, it is possible to further classify it based on how fast it happened. In this paper, we have defined two types of change: short- or long-term changes. Long-term changes occur gradually, meaning that the similarity measure between temporally close frames may large, but it decreases when more distant frames are evaluated. For instance, if a group is walking towards the same direction and the members in front of the crowd start running (and then the members behind, progressively), the behavior change will be gradual (long-term). On the other hand, short-term changes occur more abruptly, as in a panic situation in which all members of the crowd start running suddenly to different directions. Fig. 4 shows some representative frames of PETS2009 S3.Event Recognition dataset (sequence 1), in which people start running into the same direction in a progressive (long-term) manner. Fig. 5 shows a few frames of the PETS2009 S3.Event Recognition dataset (sequence 4), related to a short-term change (people suddenly start running into different directions).

In long-term changes, vector $\boldsymbol{S}_t$ tends to present values that smoothly decrease (as older frames are used in the comparison). In the second case, most values of $\boldsymbol{S}_t$ tend to be lower, since sudden changes will also affect the similarity value of recent frames. This behavior is captured by computing the first-order differences of $\boldsymbol{S}_t$ (in absolute value), stored in another vector $\boldsymbol{y}_t$. When a long-term change happens, values of $\boldsymbol{y}_t$ tend to be roughly similar, so that the maximum value should be close to the average. On the other hand, during short-term changes, the sharp drop in $\boldsymbol{S}_t$ should lead to a relatively large value of $\boldsymbol{y}_t$, but the average of $\boldsymbol{y}_t$ should be low. Hence, a change is classified as short-term if

$$c_t = \frac{\max\{\boldsymbol{y}_t\}}{\overline{\boldsymbol{y}_t}} > \alpha, \tag{7}$$

where α is a threshold set experimentally to 3.7.

[1] http://www.cvg.rdg.ac.uk/PETS2009



(a) (b)

(c)

Fig. 3. Top left: first frame of the video. Top right: frame in which the change of crowd behavior occurred. Bottom: Plot of σ_t as function of time, with a vertical red line marking the detected change.

IV. Experiments and Results

To evaluate the method presented in this paper, we used in the experiments two public datasets. The PETS2009 S3 dataset for crowd analysis and the public dataset of the University of Minnesota [23] for escape panic scenarios were used to validate the proposed approach, and the goal is to detect

(a) frame 47 (b) frame 66 (c) frame 71 (d) frame 78 (e) frame 90

Fig. 4. Frames illustrating a long-term change: people are moving to the right, and start progressively to run (first people in front, and then the others).

(a) frame 330 (b) frame 337 (c) frame 343 (d) frame 347 (e) frame 359

Fig. 5. Frames illustrating a short-term change: people are standing in the middle of the scene and suddenly start running.

changes as soon as they happen. The PETS2009 dataset provides the camera parameters, and for the Minnesota dataset we have estimated the ground plane homography using geometric planar structures present in the video sequence.

Although the Minnesota dataset presents ground truth values for event detection, they seem to be marked a few frames after the event has occurred, as noticed in [13]. Hence, we have used the updated ground truth values proposed in [13] for this database, and used the same technique to find the actual ground truth of the others sequence, as can be seen in Fig. 6.

Given the goal of detecting the change in the crowd behavior as early it occurs, in Fig. 7, it can be seen that our method presents better results than the social force model (SFM) [5] and the adjacency-matrix based clustering (AMC) [13], meaning that the event (change) was detected faster.

Fig. 8 illustrates our results for the PETS2009 S3.Event Recognition dataset – sequence 1. Ground truth was observed by us, using the same strategy adopted in [13], i.e., to manually label the frame in which the behavior change happens (in this case, when the first people start to run). Our approach detected a change 10 frames after the ground truth value, corresponding to approximately 1.5 seconds. This happens because the change occurs in a long-term fashion: people in the front of the crowd start running first, and then the others, successively. In fact, our method detected this event as a long-term change using Equation (7), and the corresponding value of c_t was 3.52. For the sake of comparison, the scenes shown in Fig. 9 and Fig. 7 are short-term changes, and presented as result $c_t = 3.91$ and $c_t = 4.35$ respectively.

In Fig. 8, Fig. 9 and Fig. 10 our method is compared with the approach developed by Briassouli and Kompatsiaris [14], in all the cases, the method presented in this paper detected the changes in crowd behavior earlier than the method compared. Despite the difference in detection time between both methods

(a) (b) (c)

(d)

Fig. 7. (a) The frame of the sequence. Event starts at frame 570, indicated in (b). Our method detects the change in crowd behavior at frame 572, shown in (c), outperforming both the social force model [5] and the adjacency-matrix based clustering approach [13], which detect the events at frames 594 and 575, respectively, as shown in (d).

be low (usually less than a second), our method is slightly better than the other in question time (in frames) of the change detection. In Fig. 10, the change is classified as a long-term change, and the detection occurred about 3 seconds after the manual annotation. As already explained before, long-term changes present a larger detection lag, due to the smooth change of σ_t.

Experiment was also conducted in other sequence of dataset Minnesota, as exemplified in Fig. 11, but with no comparison with other methods[2]. In Fig. 11, a short-term change was detected by our classifier 12 frames after the ground truth annotation.

[2]Results related to these sequences are not reported in the respective papers.

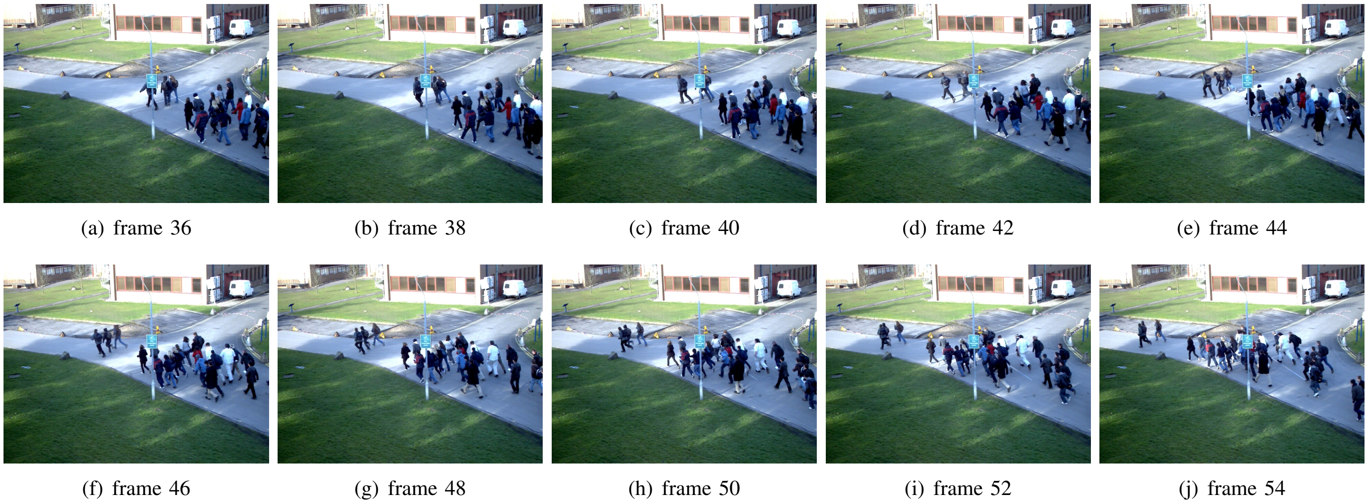


Fig. 6. Frames illustrating the detection of a short-time event. People start to run at frame 38 (b) (this value was used as ground truth), and our approach indicated the change at frame 48 (g).

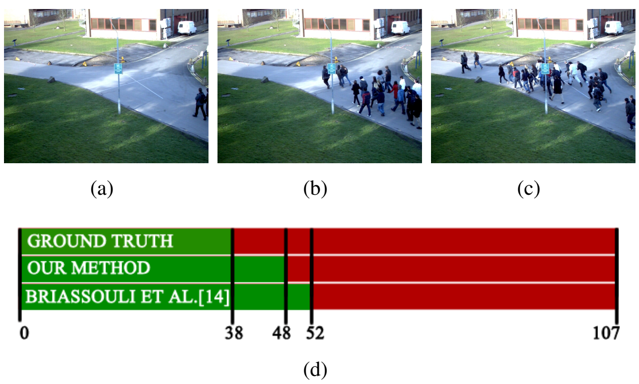


Fig. 8. (a) The first frame of the sequence. (b) The frame where the ground truth indicates a change (frame 38). (c) Frame in which our method detected the change in crowd behavior (frame 48). (d) Schematic illustration of the video sequence timeline with the ground truth and detection frames using our approach (frame 48) and the method presented in [14] (frame 52).

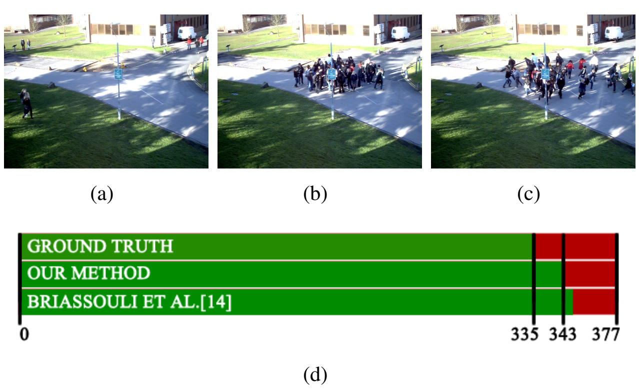


Fig. 9. (a) First frame of the analyzed sequence. (b) Frame 335, when event starts. (c) Frame 343, when our method detects the event (1.14 seconds after it started). (d) Schematic illustration of the video sequence timeline with the ground truth and detection frames using our method (frame 343) and the technique presented in [14] (frame 348).

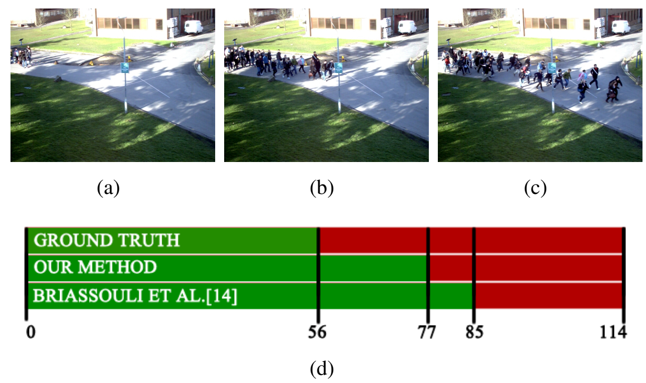


Fig. 10. (a) First frame of the sequence. (b) Frame when the ground truth indicates a change (frame 56). (c) which our method detected the event (frame 77). (d) Schematic illustration of the video sequence timeline with the ground truth and detection frames, and shown the difference between the detection using our method (frame 77) and when is used the method presented in [14] (frame 85).

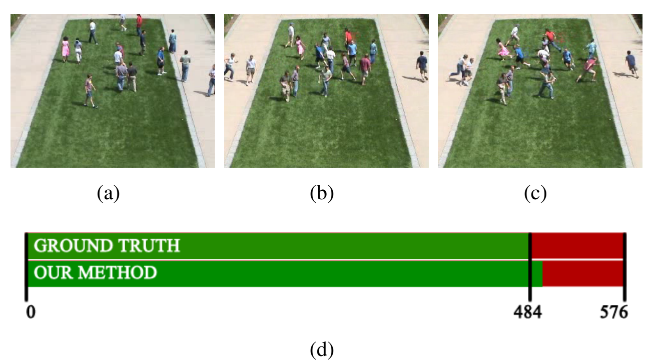


Fig. 11. (a) The first frame of the sequence. (b) The frame when event starts (frame 484). (c) Frame in which our method detected the abnormal behavior (frame 496). (d) Schematic illustration of the video sequence timeline with the ground truth value and our result.

V. Conclusion

In this paper we presented an approach to detect behavior changes in crowded scenes. The proposed method is based on the extraction of foreground blobs with shadow suppression to identify crowd members, and large-scale optical flow to obtain the displacement vector field. This vector field is mapped to world coordinates, and 2D histograms decoupling speed and orientations are computed. The similarity of motion histograms across several frames is the used to detect changes in the crowd behavior, and to classify them as short-term or long-term.

Experimental results with publicly available datasets indicate that the proposed approach can effectively detect behavioral changes, presenting an accuracy equivalent (or better than) existing approaches. Although the need of a calibrated camera may be a drawback of the proposed approach, it is important to note that there are self-calibration algorithms for obtaining the ground plane homography [20], and semi-automatic methods are becoming popular [24]. In any case, the exact same procedure may be used with the displacement vector field in image coordinates (with the manual definition of speed bin sizes). The result without calibrated camera tends to be worse, but in some cases can be useful.

As future work, we intend to further investigate how the similarity vector $\boldsymbol{S}_t$ can be explored for change detection. It is also possible to build 1D histograms of speed and orientation only, respectively, from $H_t(i,j)$. With these histograms, one can categorize if the change was mainly due to speed variations or orientation changes.

Acknowledgment

This work was partially supported by Brazilian agencies Capes and CNPq.

References

[1] G. Lavee, E. Rivlin, and M. Rudzsky, "Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video," *IEEE Transactions on Systems, Man Cybernetics - Part C*, vol. 39, no. 5, pp. 489–504, 2009.

[2] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, "Crowd analysis: a survey," *Machine Vision and Applications*, no. 19, pp. 345–357, 2008.

[3] H. Benesch, *Atlas de la Psychologie*. France: Encyclopédies d'Aujourd'hui, 1995.

[4] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6.

[5] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 935–942.

[6] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2064 –2070, oct. 2012.

[7] N. Ihaddadene and C. Djeraba, "Real-time crowd motion analysis," in *ICPR 2008*, 2008, pp. 1–4.

[8] H. M. Dee and A. Caplier, "Crowd behaviour analysis using histograms of motion direction," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 1545–1548.

[9] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, pp. 593–600.

[10] W. Li, J.-H. Ruan, and H.-A. Zhao, "Crowd movement segmentation using velocity field histogram curve," in *International Conference on Wavelet Analysis and Pattern Recognition*, 2012, pp. 191–195.

[11] T. Cao, X. Wu, J. Guo, S. Yu, and Y. Xu, "Abnormal crowd motion analysis," in *Proceedings of the 2009 international conference on Robotics and biomimetics*. IEEE Press, 2009, pp. 1709–1714.

[12] G. J. Brostow and R. Cipolla, "Unsupervised bayesian detection of independent motion in crowds," in *IEEE Computer Vision and Pattern Recognition*, 2006, pp. 594–601.

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

[14] A. Briassouli and I. Kompatsiaris, "Spatiotemporally localized new event detection in crowds." in *IEEE International Conference on Computer Vision (ICCV Workshops)*. IEEE, 2011, pp. 928–933.

[15] M. Haque and M. M. Murshed, "Panic-driven event detection from surveillance video stream without track and motion features." in *IEEE International Conference on Multimedia and Expo*, 2010, pp. 173–178.

[16] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2054–2060.

[17] M. Andersson, F. Gustafsson, L. St-Laurent, and D. Prevost, "Recognition of anomalous motion patterns in urban surveillance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 102–110, 2013.

[18] C. R. Jung, "Efficient background subtraction and shadow removal for monochromatic video sequences," *IEEE Transactions on Multimedia*, vol. 30, no. 8, June 2009.

[19] T. Brox and J. Malik, "Large displacement optical flow: descriptor matching in variational motion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 500–513, 2011.

[20] B. Bose and E. Grimson, "Ground plane rectification by tracking moving objects," in *IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, 2003, pp. 94–101.

[21] A. Rotstein, O. Inbar, T. Berginsky, and Y. Meckel, "Preferred transition speed between walking dand running: Effects of training status," *Medicine and science in sports and exercise*, no. 37, pp. 1864–1870, 2005.

[22] J. Hwang, S. Lay, and A. Lippman, "Nonparametric multivariate density estimation: a comparative study," vol. 42, no. 10, pp. 2795–2810, October 1994.

[23] U. of Minnesota Crowd Activity Dataset. http://www.mha.cs.umn.edu/movies/crowd-activity-all.avi. [Online]. Available: http://www.mha.cs.umn.edu/Movies/Crowd-Activity-All.avi

[24] Z. Zhang, T. Tan, K. Huang, and Y. Wang, "Practical camera calibration from moving objects for traffic scene surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, p. 1, 2012, available online.