Real-time Crowd Motion Analysis

Nacim Ihaddadene and Chabane Djeraba
Computer Science Laboratory of Lille (LIFL)
University of Sciences and Technologies of Lille - France
nacim.ihaddadene@lifl.fr, chabane.djeraba@lifl.fr

Abstract

Video-surveillance systems are becoming more and more autonomous in the detection and the reporting of abnormal events. In this context, this paper presents an approach to detect abnormal situations in crowded scenes by analyzing the motion aspect instead of tracking subjects one by one. The proposed approach estimates sudden changes and abnormal motion variations of a set of points of interest (POI). The number of tracked POIs is reduced using a mask that corresponds to hot areas of the built motion heat map. The approach detects events where local motion variation is important compared to previous events. Optical flow techniques are used to extract information such as density, direction and velocity. To demonstrate the interest of the approach, we present the results on the detection of collapsing events in real videos of airport escalator exits.

1. Introduction

Public safety is increasingly a major problem in public areas such as airports, malls, subway stations, etc. Processing the recorded videos can be exploited to present informative data to the security team who needs to take prompt actions in a critical situation and react in case of unusual events. In the last years, most of surveillance systems integrated computer vision algorithms to deal with problems like motion detection and tracking. However, a few have treated the problems involving crowded scenes due to the problematic complexity.

Recently, the computer vision community has started taking interest in addressing different research problems related to the scenarios involving large crowds of people. The focus so far has been on the tasks of crowd detection and tracking of individuals in the crowd.

Our approach is based on motion variation on regions of interests. First, a motion intensity heat map is computed during a certain period of time. The motion heat map represents hot and cold areas on the basis of motion intensities. The hot areas are the zone of the scene where the motion is high. The cold areas are regions of the scene where the motion intensities are low. More generally, we consider several levels of motion intensity. The points of interest are extracted in the selected regions of the scene. Then, these points of interest are tracked using optical flow techniques. Finally, variations of motions are estimated to discriminate potential abnormal events. The main advantage of the approach is the fact that it doesn't require a huge amount of data to enable supervised/unsupervised learning. It detects all events where the variations are important.

The structure of this paper will start by presenting some related works in the next section. Sections 3 and 4 will discuss the different steps to estimate the abnormality of a crowd flow. In section 5, a case of processing algorithm application is presented with an example of an escalator camera. Finally, section 6 presents the conclusion and some future extensions of the work.

2. Related Works

There are two categories of related works. The first category is related to crowd flow analysis, and the second category is related to abnormal event detection in crowd flows. The works of the first category, estimate crowd density [12, 13, 9, 11]. These methods are based on textures and motion area ratio and make an interesting static analysis for crowd surveillance, but do not detect abnormal situations. There are also some optical flow based techniques [4, 6] to detect stationary crowds, or tracking techniques by using multiple cameras[5].

The works in the second category detects abnormal events in crowd flows. The general approach consists of modelling normal behaviors, and then estimating the deviations between the normal behavior model and the

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

observed behaviors. These deviations are labelled as abnormal.

The principle of the general approach is to exploit the fact that data of normal behaviors are generally available, and data of abnormal behaviors are generally less available. That is why, the deviations from examples of normal behavior are used to characterize abnormality. In this category, [2, 3] combines HMM, spectral clustering and principal component for detecting crowd emergency scenarios. The method was experimented in simulated data. [1] uses Lagrangian Particle Dynamics for the detection of flow instabilities, this method is efficient for segmentation of high density crowd flows (marathons, political events, ...).

Our approach contributes to the detection of abnormal events in crowd flows by the fact that it doesn't need learning process and training data, it also considers simultaneously density, direction and velocity and focuses analysis on specific regions where the density of motions is high.

3 Algorithm steps

The proposed algorithm is composed of several steps: motion heat map and direction map building, features extraction, optical flow calculation and finally the estimation of abnormality description measure.

3.1 Motion heat map

Motion heat map is a 2D histogram indicating the main regions of motion activity. This histogram is built from the accumulation of binary blobs of moving objects, which were extracted following background subtraction method [8].

The obtained map is used as a mask to define the Region of Interest (RoI) for the next step of the processing algorithm such as "Features detection" (described later). The use of heat map image improves the quality of the results and reduces processing time which is an important factor for real-time applications. Figure 1 shows an example of the obtained heat map from an escalator camera view. The results are more significant when the video duration is long. In practice, even for the same place, the properties of unusual events may vary depending on the context (day/night, indoor/outdoor, normal/peak time, ...). We built a motion heat map for each set of conditions.

3.2 Features detection and tracking

In this step, we start by extracting a set of points of interest from each input frame. We use a mask to define

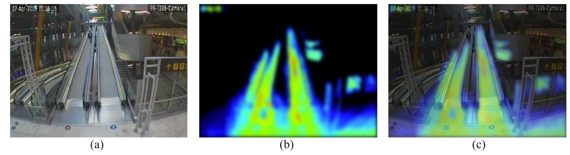

Figure 1. Example of motion heat map generation. (a) camera view, (b) generated motion heat map and (c) masked view.

a region of interest. This mask is obtained from the hot areas of the heat map image. In our approach, we consider Harris corner as a point of interest [7]. We consider that in video surveillance scenes, camera positions and lighting conditions allow to get a large number of corner features that can be easily captured and tracked.

Once we define the points of interest, we track these points over the next frames using optical flow techniques. For this, we used Kanade-Lucas-Tomasi feature tracker [10, 14]. After matching features between frames, the result is a set of vectors:

$$V = \{V_1 ... V_N | V_i = (X_i, Y_i, A_i, M_i)\}$$

where X_i and Y_i are the coordinates of the feature i, A_i is the motion direction of feature i and M_i is the distance between the feature i in the frame t and its matched feature in frame $t + 1$.

This step also allows removal of static and noise features. Static features are the features that moves less than two pixels. Noise features are the isolated features that have a big angle and magnitude difference with their near neighbors due to tracking calculation errors.

Images in figure 2 show the set of vectors obtained by optical flow feature tracking in two different situations. The left image shows an organized vector flow. The right one shows a cluttered vector flow due to the collapsing situation.

Figure 2. Example of vector flows.

4 Direction map

Our actual approach considers also the "Direction Map" which indicates the average motion direction for each region of the camera field. The camera view is divided into small blocks and each block is represented by the mean motion vector. The distance between average direction histogram and instant histogram corresponding to the current frame increases in case of collapsing situations.

The right image of Figure 3 shows the mean direction in each block of the view field of an escalator camera. Some tendencies can be seen. In the blue region, the motion is from top to bottom. In the yellow region, the motion is from right to left.

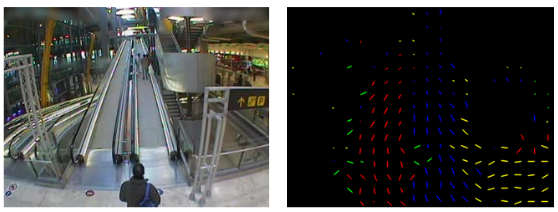

Figure 3. Example of block direction histogram : top to bottom for the left escalator, bottom to top for the right one.

4.1 Measuring entropy

In this step, we define a statistic measure that will describe how much the optical flow vectors are organized or cluttered in the frame. We studied a set of statistical and optical measures (variance, entropy, heterogeneity and saliency). The result is a measure M which is the scalar product of the normalized values of the following factors:

Motion area ratio: In crowded scenes the area covered by the moving blobs is important compared to uncrowded scenes. This measure is also used in density estimation techniques.

Direction variance: After calculating the mean direction of the optical flow vectors in a frame, we calculate the direction variance of these vectors.

Motion magnitude variance: Observation shows that this variance increases in abnormal situations. With one or many people walking even in different directions, they tend to have the same speed; which means a small value of the motion magnitude variance. It's not the case in collapsing situations and panic behaviors that often engender a big value for the motion magnitude variance.

Direction histogram peaks: The calculation of vectors direction and magnitude variances is not sufficient. We build a direction histogram in which each column indicates the number of vectors in a given angle. The result is a histogram that indicates the direction tendencies, and the number of peaks in this histogram represents the different directions.

4.2 Deciding: Normal / Abnormal

This decision is taken in a "Static way" by comparing the calculated and normalized measure with a specific threshold. Configuration has been necessary to estimate the Normal/Abnormal threshold because it varies depending on camera position, escalator type and position. The decision may also be taken in a "Dynamic way" by detecting considerable sudden changes of the cluttering measure through time.

5 Experimental Results

In our experiments, we used a set of real videos provided by cameras installed in an airport to monitor the situation of escalator exits. Videos are exploited to present informative data to the security team who needs to take prompt actions in the critical situation of collapsing.

The data set is divided into two kind of situations: normal situations and abnormal situations, with a subset of 20 videos for each type. The normal situations correspond to crowd flows without collapsing in the escalator exits. Generally, in the videos, we have two escalators, corresponding to two traffic ways, in opposite directions. Abnormal situations correspond to videos that contain collapsing events in escalator exists.

The original video frame size is 640x480 pixels and each video sequence for the different situations has more than 4000 frames. For the features detection and tracking we extract about 1500 features per frame (including static and noise features).

Figure 4 shows an example of a collapsing situation in an escalator exit. The variation of the cluttering measure M through time is shown in the graph. The red part of the curve represents the time interval where the collapsing event happened.

From the video data set, our approach detected all collapsing events. The collapsing events, detected by the system, have been compared with collapsing events annotated manually. Till now, the result is very satisfactory. However, it is necessary to select the appropriate threshold and the regions of interest carefully.

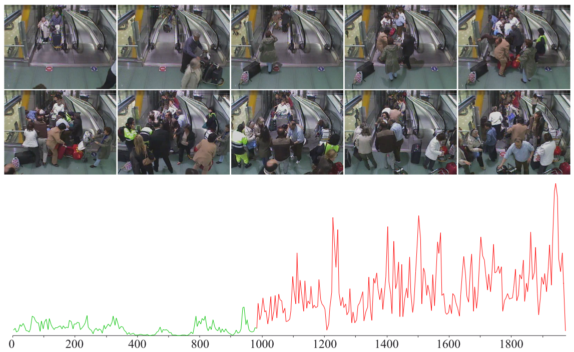

Figure 4. Measure variation.

6 Conclusion

In this paper, we proposed a method that estimates the abnormality of a crowd flow. We defined a measure that is sensitive to crowd density, velocity and direction. The method doesn't require prior specific training processes and it has been applied to detect collapsing events in airport escalator exits. The results till now are promising on it's robustness.

The work is still in progress, it is expected to extend the estimation of the motion variations with factors such as acceleration by tracking the POIs over multiple frames. The contextualization of the system is also important. Introducing context information to optimize the system configuration allows to use the same underlying detection algorithms in different locations and in an efficient way. The understanding of the system configuration is also important for the security team to assess the current situation.

Acknowledgements. This work has been performed by members of MIAUCE project, which is a 6^{th} Framework Research Programme of the European Union (IST-2005-5-033715), The authors would like to thank the EU for the financial support and the partners within the project for a fruitful collaboration. For more information about the MIAUCE project, visit: http://www.miauce.org.

References

[1] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–6, 17-22 June 2007.

[2] E. Andrade, S. Blunsden, and R. Fisher. Hidden markov models for optical flow analysis in crowds. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference*, vol. 1:460–463, 2006.

[3] E. L. Andrade, S. Blunsden, and R. B. Fisher. Modelling crowd scenes for event detection. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 175–178, Washington, DC, USA, 2006. IEEE Computer Society.

[4] B. Boghossian and S. Velastin. Motion-based machine vision techniques for the management of large crowds. *Electronics, Circuits and Systems, 1999. Proceedings of ICECS '99. The 6th IEEE International Conference*, vol. 2:961–964, 1999.

[5] F. Cupillard, A. Avanzi, F. Bremond, and M. Thonnat. Video understanding for metro surveillance. *Networking, Sensing and Control, 2004 IEEE International Conference*, vol. 1:186–191, 2004.

[6] A. Davies, J. H. Yin, and S. Velastin. Crowd monitoring using image processing. *Electronics & Communication Engineering Journal*, vol. 7(1):37–47, 1995.

[7] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, page 147152, 1988.

[8] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian. Foreground object detection from videos containing complex background. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 2–10, New York, NY, USA, 2003. ACM.

[9] S.-F. Lin, J.-Y. Chen, and H.-X. Chao. Estimation of number of people in crowded scenes using perspective transformation. *Systems, Man and Cybernetics, Part A, IEEE Transactions*, 31(6):645–654, 2001.

[10] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proceedings of the International Joint Conference on Artificial Intelligence*, (1):pp 674–679, 1981.

[11] R. Ma, L. Li, W. Huang, and Q. Tian. On pixel count based crowd density estimation for visual surveillance. *Cybernetics and Intelligent Systems, 2004 IEEE Conference*, vol. 1:170–173, 2004.

[12] A. Marana, S. Velastin, L. Costa, and R. Lotufo. Estimation of crowd density using image processing. *Image Processing for Security Applications (Digest No.: 1997/074), IEE Colloquium*, pages 11/1–11/8, 1997.

[13] H. Rahmalan, M. S. Nixon, and J. N. Carter. On crowd density estimation for surveillance. In *International Conference on Crime Detection and Prevention*, 2006.

[14] J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.