

Motion modeling for abnormal event detection in crowd scenes

Hedi Tabia, Michèle Gouiffes and Lionel Lacassagne
IEF, Institut d'Electronique Fondamentale
Bat 220, Campus Scientifique d'Orsay
91405 Orsay, FRANCE

Abstract—Since few years, video surveillance systems have become quite popular and have been undertaken in order to enhance our sense of security. In this context, this paper presents an approach for abnormal situation detection in video scenes. By analyzing the motion aspect of low level features instead of tracking segmented objects one by one, the proposed approach aims to produce a probabilistic model for normal situations. Abnormal ones are detected by comparing the real-time observation with usual situation parameters. The probabilistic model estimates abrupt changes and abnormal behaviors of a set of pixels selected from each frame in the scene. To demonstrate the interest of the methodology, we present the obtained results on videos that contain running of the bulls scenes, violence in areas where quiet must prevail and some collapsing events in real videos amassed by single camera.

I. INTRODUCTION

As the safety of people and their goods is a major problem in public areas such as airports, subway stations, shopping centers or any other public places, the use of video surveillance has grown very rapidly. Most existing video surveillance systems are fully supervised by humans resulting in poor performance. The computer vision and the multimedia computing communities have recently witnessed a surge of methods that contribute to the automatic video surveillance. Of all these methods, few have addressed the abnormal event detection problem in crowd scenes. This is due to the complexity of the task when dealing with generic situations or situations involving crowds of people. In video surveillance applications, abnormal events can be described as unusual events encountered at a specific context and may be reported for further investigations.

A. Related work

Through the state-of-the-art, we can find mainly three categories of related work. The first category deals with the estimation of the crowd density, the second category is applied to abnormal event detection in crowd flows, and the third category is associated to the indexing of abnormal events.

In the first category, several authors like [1], [2], [3] have been interested by the estimation of crowd density. These works are generally 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], [5] to detect stationary crowds, or tracking techniques by using multiple cameras [6].

The general approach in the second category of the related work, consists in modeling normal behaviors, and then estimating the deviations between the normal behavior model and the observed behaviors. These deviations are labeled 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, Andrade and Fisher [7] have presented an event detector for emergencies in crowds. Authors encode the optical flow features with Hidden Markov Models to allow for the detection of crowd emergency scenarios. In order to increase the detection sensitivity a local modeling approach is used. The method was experimented on simulated data. Saad and Mubarak [8] proposed a framework in which Lagrangian Particle Dynamics is used for the segmentation of high density crowd flows and detection of flow instabilities. Their method is efficient for segmentation of high density crowd flows (marathons, political events, ...).

Compared to the work on the estimation of crowd density, the state of the art concerning the indexing of abnormal events is relatively poor. Xie et al. [9] proposed a semantic based retrieval framework for traffic video sequences. In order to estimate the low-level motion data, a cluster tracking algorithm is developed. A hierarchical self-organizing map is applied to learn the activity patterns. By using activity pattern analysis and semantic concepts assignment, a set of activity models (to represent the general and abnormal behaviors of vehicles) are generated, which are used as the indexing-key for accessing video clips and individual vehicles in the semantic level. The proposed retrieval framework supports various queries including query by keywords, query by sketch and multiple object queries.

B. The approach

As shown in Figure 1, the proposed approach consists of four main steps. Following the first (background subtraction) and the second one (moving connected component detection), we aim to extract a set of interest pixels from the current frame. These pixels are then tracked in the following frames using optical flow. The fourth step intends to model the scene frames using a probabilistic model that codes the normal situations and thus allows the system to pick out abnormal cases. The model can be updated over all the frames of the scene. In

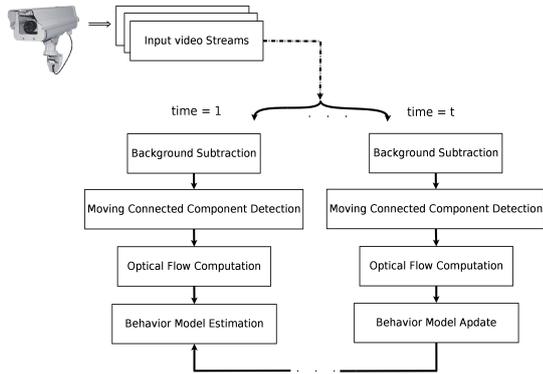


Fig. 1. Method overview

our work the modeling of the frames is performed pixel by pixel. The idea has some analogy with ideas developed by Stauffer and Grimson [10] for background subtraction. The approach we propose, consists in studying the behavior of each pixel in the frame during some laps of time called the *pixel behavior history*. A pixel from a new frame is considered to have a normal behavior if its new value is well described by its density function. In the contrary, an abnormal behavior is obtained when the current observation does not match with the pixel behavior history. The pixel values used in this work are based on the magnitude and the direction of the optical flows.

The rest of this paper is organized as follows. First, section II presents the motion detection and the motion tracking steps. Section III explain the modeling of the scene behaviors. Experiments are conducted in section IV. Section V concludes the paper.

II. MOTION DETECTION AND TRACKING

Moving object detection and tracking are common computer vision tasks. The moving object detection task aims to segment the regions that contain moving pixels from the rest of the image. The object tracking task consists in estimating the location of each object in each new frame. These two tasks can either be performed separately or jointly [11]. In the first case, possible object regions in every frame are obtained by means of an object detection algorithm, and then the tracker corresponds objects across frames. In the second case, the object region and correspondence is jointly estimated by iteratively updating object location and region information obtained from previous frames. In our work, these tasks are performed separately in two steps

A. Moving region detection

Moving object detection is a crucial step in our analysis since the detection of abnormal events is greatly dependent on the performance of this stage. The purpose of this step is to improve the quality of the results while reducing processing time by considering only moving regions. Many algorithms have been suggested to solve the problem of moving object detection. A generally applicable hypothesis in video

surveillance is that the images of the scene without the presence of moving objects present an uniform behavior which can be modeled by a statistical model usually a probability distribution. So that given the statistical model of the scene, a moving object can be detected easily by identifying parts of the image that do not match with the model. This process is known as *background subtraction*. In our work, we use an implementation of this process proposed in [12]. Due to inaccurate results of background subtraction process, it is necessary to refine the detection by applying a segmentation process, which removes the small and isolated noise while extracting the most significant moving objects. Note that, in our analysis we are more interested in moving pixels rather than moving objects, but we focus on the pixels belonging to significant moving objects.

B. Moving region tracking

By locating the position of the interest pixels in every frame of the video, the aim of this step is to compute the optical flow hence to approximate the 2D motion field. In our surveillance system, we use a tracking algorithm based on the Kanade-Lucas-Tomasi feature tracking [13], [14]. After matching features between frames, the result is a set of vectors:

$$V_t = \{v_1, \dots, v_N | v_i = (A_i, D_i)\} \quad (1)$$

where i denotes the feature pixel located in (x_i, y_i) , $A_i = (\cos(\phi_i), \sin(\phi_i))$ represents the Cosine and the Sine of the motion direction ϕ_i of the feature i and D_i is the distance between the feature i in the frame t and its matched feature in the frame $t + 1$. Figure 2 shows an example of motion detection and tracking results. Figure 2 (a) presents a frame of a video sequence showing running people in a normal situation. Figure 2 (b) shows a motion heat map indicating the main regions of motion activity. This map is built from the accumulation of binary blobs of moving objects, which were extracted following the background subtraction method. Figure 2 (c) shows the disjoint objects detected after the segmentation step. Figure 2 (d) presents the optical flows computed on each pixel in the moving areas. The colors in this figure indicate the direction of the optical flow vectors. In this visualization, we code only two directions of the optical flow. The black color refers to pixels with insignificant motion magnitude.

III. SCENE BEHAVIOR MODEL

In this section, we assume that pixels in the frames of the scene with normal situation exhibit some regular behavior which can be described by a probability density function. If the statistical model of the scene is available, the abnormal situation can be detected by minimizing the likelihood with the normal situation. To this end, we have developed an efficient algorithm inspired from the work of Stauffer and Grimson [10] for background subtraction. Our pixel-based method can be described as follows. We consider the vector V_t of eq (1) which collects the characteristics of a pixel at time t . The probability density function modeling a pixel is noted $p(V_t)$. The pixel is labeled as *Normal* if its new value V_{t+1} is

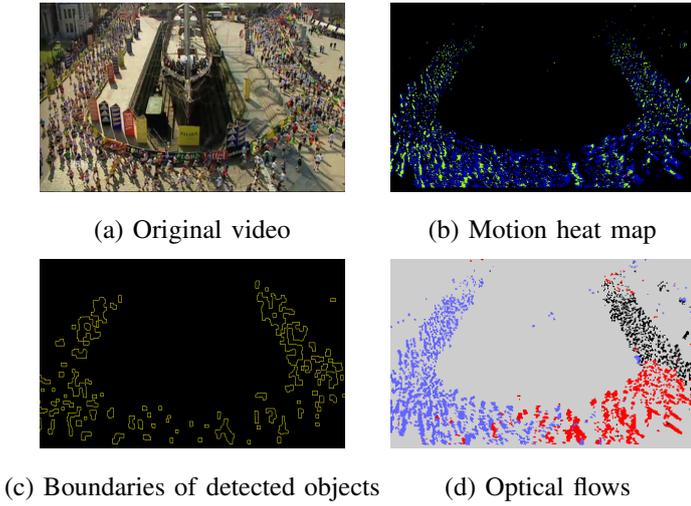


Fig. 2. Motion detection and tracking processes.

well described by its density function and abnormal otherwise. Formally, at any time, t , what is known about a particular pixel at location, $\{x_i, y_i\}$, is its history. The recent history of each pixel, $\{V_1, \dots, V_t\}$, is modeled by a mixture of M Gaussian distributions. The probability of observing the current pixel value is:

$$p(V_t; \theta(M)) = \sum_{m=1}^M \omega_m p_m(V_t; \theta_m) \quad \text{with} \quad \sum_{m=1}^M \omega_m = 1 \quad (2)$$

where $\theta(M) = \{\omega_1, \dots, \omega_M, \theta_1, \dots, \theta_M\}$ and $p_m(V_t; \theta_m)$ is the normal distribution of the m^{th} component represented by:

$$p_m(V_t; \theta_m) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma_m|^{\frac{1}{2}}} e^{-\frac{1}{2}(V_t - \mu_m)^T \Sigma_m^{-1} (V_t - \mu_m)} \quad (3)$$

where $\Sigma_m = \sigma_m^2 I$ is the covariance of the m^{th} component of the Gaussian mixture and I refers to the identity matrix. Every new pixel vector V_t is compared with the existing M Gaussian distributions, until a match is found. If none of the M distributions match the current point vector, the least probable distribution is replaced with a distribution with the current vector as its mean vector, an initially high variance, and low prior weight. At the instant t , the different weights of the M Gaussians, are updated according to this formula:

$$\omega_{m,t} = (1 - \alpha)\omega_{m,t-1} + \alpha(G_{m,t}) \quad (4)$$

Here α refers to the learning rate and $G_{m,t}$ is set to be 1 if the pixel value is generated by the Gaussian m and 0 otherwise. The mean and the variance of the Gaussian which fit the new observation are updated according to:

$$\begin{aligned} \mu_t &= (1 - \rho)\mu_{t-1} + \rho V_t \\ \sigma_t^2 &= (1 - \rho)\sigma_{t-1}^2 + \rho(V_t - \mu_t)^T (V_t - \mu_t) \end{aligned} \quad (5)$$

where $\rho = \alpha p_m(V_t; \theta_m)$. As we mentioned earlier this statistical process aims to build a compact model for the behavior of the scene. Although this model is updated regularly according to eq (4) and eq (5) which make the model more flexible to the behavior changes, the user can stop the update process by setting $\alpha = 0$. This can be applied when monitoring scenes provided by static cameras installed in airport issues or subway stations for example. In this case, it is important to have long scene duration after which, we stop the update process.

A. Normal situation model

In accordance with the reasoning presented in the previous section, the mixture of Gaussians models both situations the normal and the abnormal one. In order to pick out only the distributions modeling the normal situation, the components of mixture of each pixel are ranked based on the value of the quotient $\frac{\omega_m}{\sigma_m}$ and only the first $N < M$ distributions are taken as normal situation model [10].

$$N = \underset{n}{\operatorname{argmin}} \left(\sum_{m=1}^n w_m > Th \right) \quad (6)$$

where Th corresponds to the portion of the data which has to be accounted by the normal situation.

B. Abnormal Connected components

The probabilistic model that we presented above is constructed for each moving pixel in the scene. We remind here that these pixels are obtained following a segmentation process (see section II-A). One of the significant advantages of this process is that when we are looking for abnormal behaviors of a whole object, it is easy to consider only connected components which contain abnormal pixel behaviors.

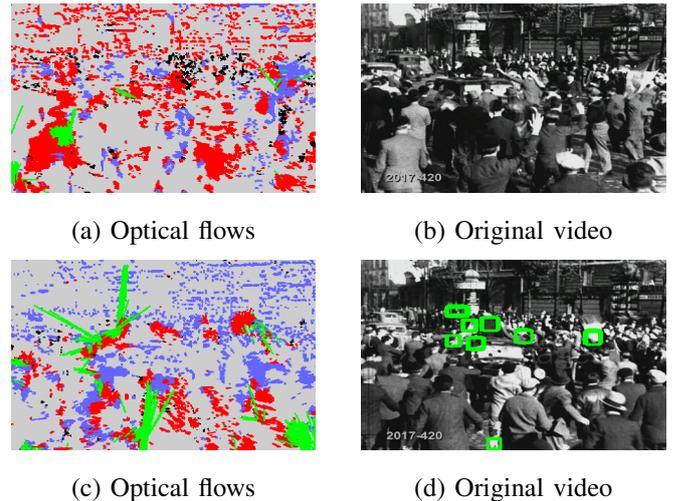


Fig. 3. An example of pixel behaviors in abnormal situation.

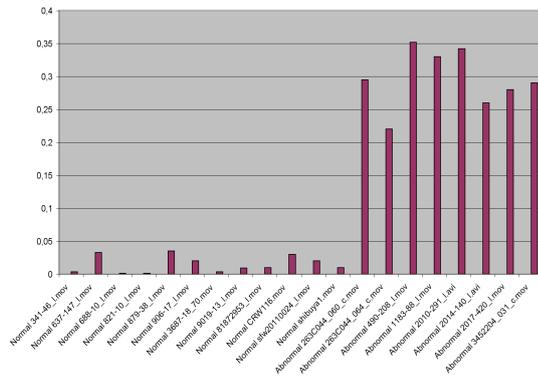


Fig. 4. The rate of abnormal pixels returned by our algorithm over all moving pixels in the different video sequences..

IV. EXPERIMENTAL RESULTS

We demonstrate the performance of our approach using a set of videos¹ proposed by Mehran et al. [15]. The data set is composed of 20 videos arranged into two kinds of situations: normal situations with 12 videos and abnormal situations with 8 videos. The normal situations correspond to generic crowd flows. Abnormal situations correspond to videos that contain running of the bulls videos, violence in areas where quiet must prevail and some collapsing events. The frame sizes in both situations are different. Figure 3 shows an example of an abnormal situation. Figure 3 (a) shows the optical flows computed in moving pixels. Figure 3 (b) shows the output of our algorithm corresponding to the same frame in Figure 3 (a). We notice here that the algorithm does not detect any abnormal behavior of pixel, while a considerable sudden changes in the optical flow values (Green color in Figure 3 (a)). In fact at the instant when we selected the frame from the scene, the abnormal situation which corresponds to a jostling event concentrated in the center of the frame, does not start yet. And the sudden changes in the optical flow values shown in Figure 3 (a) are due to the occlusion made by the crowds. Since this kind of moving through occlusion have been already learned from the previous frames, our algorithm does not detect any abnormal behavior of pixel. Moreover, as shown in Figure 3 (d), our algorithm detects a set of abnormal pixel behaviors and ignores the others. The set of pixels returned by our algorithm is mainly located around the central region of the frame which corresponds to the jostling location. From the video data set, our approach successfully discriminates between both kinds of events. Figure 4 shows the rate of abnormal pixels returned by our algorithm over all moving pixels in the different video sequences. We can see that the algorithm succeeds in detecting a large amount of abnormal pixels in the videos considered as "abnormal". The normal videos can also contain abnormal pixels, but the global average is not significant (under 0.05 %) and may be produced by noise.

¹The dataset can be downloaded from the author home page: <http://www.cs.ucf.edu/~ramin/>

V. CONCLUSION

In this paper, we presented an approach that detects abnormal event in crowd flow. The approach is similar in its spirit to background subtraction methods used in video surveillance systems. It consists in studying the behavior of each pixel in the frame during a time history. A pixel from a new frame is considered to have a normal behavior if its new value is well described by its density function. In the contrary, an abnormal behavior is obtained when the current observation does not match with the pixel behavior history. The pixel values used in this work are based on the magnitude and the direction of the optical flows. Although the work is still in progress, the results shows that the approach is very promising. Note also that quite all algorithms are embarrassingly parallel which make the application compatible with parallel implementation on a multi-core general purpose processor or on a GPU for real time processing. These points will be addressed soon.

ACKNOWLEDGMENT

This research is supported by the European Project ITEA2 SPY Surveillance imProved sYstem <http://www.ppsl.asso.fr/spy.php>.

REFERENCES

- [1] H. Rahmalan, M. S. Nixon, and J. N. Carter, "On crowd density estimation for surveillance," *IET Seminar Digests*, vol. 2006, no. 11434, pp. 540–545, 2006.
- [2] A. N. Marana, S. A. Velastin, L. F. Costa, and R. A. Lotufo, "Estimation of crowd density using image processing," 1997.
- [3] Ruihua Ma, Liyuan Li, Weimin Huang, and Qi Tian, "On pixel count based crowd density estimation for visual surveillance," in *IEEE Conference on Cybernetics and Intelligent Systems, 2004.*, 2005, pp. 170–173.
- [4] *Motion-based machine vision techniques for the management of large crowds*, vol. 2, 2002.
- [5] A. C. Davies, J. H. Yin, and S. Velastin, "Crowd monitoring using image processing," *Electronics and Communication Engineering Journal*, 1995.
- [6] Frederic Cupillard, Alberto Avanzi, Francois Bremond, and Monique Thonnat, "Video understanding for metro surveillance," in *IEEE International Conference on Networking, Sensing and Control*, 2004.
- [7] E. L. Andrade and R. B. Fisher, "Hidden markov models for optical flow analysis in crowds," in *Proc. International Conference on Pattern Recognition (ICPR)*, 2006, vol. 1, pp. 460–463.
- [8] *A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis*, 2007.
- [9] Dan Xie, Weiming Hu, Tieniu Tan, and Junyi Peng, *Semantic-based traffic video retrieval using activity pattern analysis*, pp. 693–696, 2004.
- [10] Chris Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2, pp. 2246, 1999.
- [11] Alper Yilmaz, Omar Javed, and Mubarak Shah, "Object tracking: A survey," 2006.
- [12] P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for realtime tracking with shadow detection," 2001.
- [13] Bruce D. Lucas and Takeo Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2*, 1981, pp. 674–679.
- [14] Jianbo Shi and Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, 1994, pp. 593–600.
- [15] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2009.