

Detection of Abnormal Motions in Video

Mahmoudi Sidi Ahmed, Md. Haidar Sharif, Nacim Ihaddadene, Chabane Djeraba
Computer Science Laboratory of Lille (LIFL)
University of Science and Technology of Lille
{mahmoudi.sidi-ahmed, md-haidar.sharif, nacim.ihaddadene, chabane.djeraba}@lifl.fr

ABSTRACT

This paper describes an approach to detect abnormal motion in videos. The core of the approach detects portion of video that corresponds to sudden changes of motion variations of a set of defined points of interest. Optical flow technique tracks those points of interest. There are sufficient variations in the optical flow patterns in a mob scene when there are cases those showing abnormalities. The geometric clustering algorithm, *k-means*, clusters the obtained optical flow information to get the distance between two consecutive frames. In general, comparatively high distance indicates abnormal motion. To demonstrate the interest of the approach, we present the results based on the detection of abnormal motions in video, which consists of both normal and abnormal motions.

Keywords

Optical flow, Abnormal motion, K-means

1. INTRODUCTION

Visual surveillance is currently one of the active research topics in computer vision. The increasing concern about public safety and law enforcement has caused a great deal of growth in the number of surveillance cameras. Due to this fact, the necessity of automatic techniques which process and analysis human behaviours and activities is more evident everyday. However, motion detection is a fundamental processing step in the majority of visual surveillance algorithms. Motion detection algorithms mainly aim to detect moving objects while suppressing the effects caused by lighting changes, moving background, shadows, etc. In video surveillance motion detection could take a vital role for security and safety in public places such as concerts, sporting events, parking places, town centers, political events, etc. The detection of abnormal motion would benefit from a system capable of recognizing perilous and inconsistent conditions and circumstances to make the system operators fully aware and attentive.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Chania ICMI-MIAUCE'08 workshop, Crete, Greece
Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

We will present an approach which is characterized by optical flow patterns of human behaviours and activities followed by some geometrical treatments to detect abnormal motion in video. The optical flow information from video presents the crowd multi-modal behaviors as optical flow patterns variate in time. There is sufficient perturbation in the optical flow pattern in the crowd in case of abnormal and/or emergencies situations [1, 2]. We will estimate the variations of motions to discriminate potential abnormal motions without mentioning any restriction on the number of people or moving objects.

The rest of this paper is organized as follows: Section 2 describes the related work; Section 3 is devoted to the presentation of the our proposed approach; Section 4 introduces the experimental results; and finally, Section 5 concludes and indicates future work.

2. RELATED WORKS

We categorize two type of related works based on the works where motion detection is the fundamental processing step: one is related to mob flow analysis, and the other is related to abnormal event detection in mob flows.

Mob flow analysis: These works estimate mob density [9, 8]. The applied methods are based on textures and motion area ratio and make an interesting static analysis for crowd surveillance, but do not detect abnormal situations.

Abnormal event detection: The general approach of these works consists of modeling normal behaviours, and then estimating the abnormal behaviour or attitudes between the normal behaviour model and the observed behaviours. Finally the variations are labeled as abnormal. The principle of the general approach is to exploit the fact that data of normal behaviours are generally available, and data of abnormal behaviours are ordinarily unavailable. For this reason, the deviations from examples of normal behaviour are used to characterize abnormality. In this category, [1, 2] combines HMM, spectral clustering and principal component for detecting crowd emergency scenarios. The method was experimented in simulated data. Authors in [12] proposed a visual monitoring system that passively observes moving objects in a site and learns patterns activity from those observations, detect of unusual events in the site that do not fit common patterns using a hierarchical classification. Also, authors in [3] address the problem of detecting irregularities in visual data as a process of constructing a puzzle: regions in the observed data which can be composed using large contiguous chunks of data from the database are considered very likely, whereas regions in the observed data

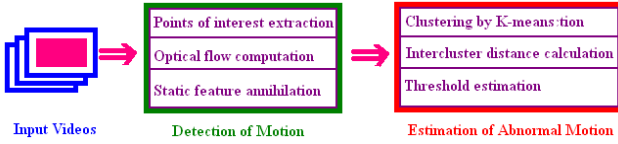


Figure 1: Block diagram of the proposed approach

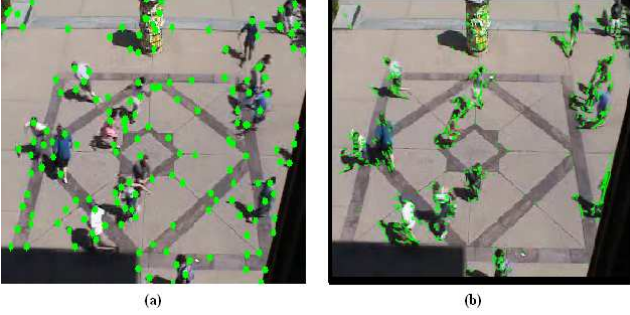


Figure 2: The green points in (a) indicate Harris points of interest, and (b) depicts optical flow.

which cannot be composed from the database are regarded as suspicious. The spatial and spatio-temporal appearance-based patch descriptors are generated for each query and for each database patch. The inference problem is posed as inference process in probabilistic graphical model.

Our approach contributes to the second category of related works possessing an advantage of without any restriction on the number of people or moving objects.

3. PROPOSED APPROACH

In this section we will discuss the detailed implementation steps of our approach. The approach consists of two steps *detection of motions* and *estimation of abnormal motions*. A block diagram of our proposed framework has been depicted in Figure 1.

3.1 Detection of Motions

3.1.1 Points of interest extraction

We consider Harris corner as a point of interest [5]. The Harris corner detector is a famous point of interest detector due to its strong invariance to rotation, scale, illumination variation, and image noise [4]. It is based on the local auto-correlation function of a signal, where the local auto-correlation function measures the local changes of the signal with patches shifted by a small amount in different directions. A discrete predecessor of the Harris detector was depicted by Moravec [10], where the discreteness refers to the shifting of the patches. Figure 2(a) shows an example of Harris points of interest.

3.1.2 Optical flow computation

Once we define the points of interest, we track these points over the next frames using optical flow techniques (see figure 2(b)). For this, we use Kanade-Lucas-Tomasi feature tracker [7, 11]. Upon matching points of interest (features) between frames, the result is a set of vectors:

$$\Omega = \{\Omega_1 \dots \Omega_N | \Omega_i = (x_i, y_i, v_i, \alpha_i)\}$$

where

x_i : x coordinate of the feature i ,

y_i : y coordinate of the feature i ,

v_i : velocity of the feature i ,

α_i : direction of motion of the feature i .

3.1.3 Static feature annihilation

This step allows removal of static and noise features. Features (points of interest) having a velocity equal to zero (i.e., $v_i = 0$) are considered as static features. Noise features are the isolated features that have a big angle and distance calculation errors. The resulting points of interest are suitable for clustering.

3.2 Estimation of abnormal motions

3.2.1 Clustering by K-means

After static error suppression points of interest, we apply K-means method to get clusters. The k-means method is a well known geometric clustering algorithm based on the work done by Lloyd in 1982 [6]. The k-means algorithm is a simple and fast method for partitioning data points into clusters. Let $X = \{x_1, x_2, \dots, x_n\}$ be a set of points in \mathbb{R}^d (d dimensional real numbers). After being seeded with a set of k centers c_1, c_2, \dots, c_k in \mathbb{R}^d , the algorithm partitions these points into clusters as follows:

1. For each $i \in \{1, \dots, k\}$, set the cluster \mathcal{C}_i to be the set of points in X that are closer to c_i than they are to c_j for all $j \neq i$.
2. For each $i \in \{1, \dots, k\}$, set c_i to be the center of mass of all points in \mathcal{C}_i : $c_i = \frac{1}{|\mathcal{C}_i|} \sum_{x_j \in \mathcal{C}_i} x_j$.
3. Repeat steps 1 and 2 until c_i and \mathcal{C}_i no longer change, at which point return the clusters \mathcal{C}_i .

If there are two centers equally close to a point in X , we break the tie arbitrarily. If a cluster has no data points at the end of step 2, we eliminate the cluster and continue as before.

3.2.2 Calculating distance between clusters

We represent each cluster as a rectangle, as shown in figure 3. We calculate the distances of all the clusters between two consecutive frames. In Figure 3 (a), assume two corners $p1$ and $q1$ of rectangles p and q possessing the coordinates $(c1x1, c1y1)$ and $(c2x2, c2y2)$ and also consider the width and height of the rectangles are $w1$ & $h1$ and $w2$ & $h2$ successively. Hence the coordinates of $p2, p3, q2,$ and $q3$ are $(c1x1 + w1, c1y1), (c1x1, c1y1 + h1), (c2x2 + w2, c2y2),$ and $(c2x2, c2y2 + h2)$ respectively. Now, we calculate the distance Γ of two rectangles as:

$$\begin{aligned} \Gamma = & \sqrt{(c1x1 - c2x2)^2 + (c1y1 - c2y2)^2} \\ & + \sqrt{(c1x1 + w1 - c2x2 - w2)^2 + (c1y1 - c2y2)^2} \\ & + \sqrt{(c1x1 - c2x2)^2 + (c1y1 + h1 - c2y2 - h2)^2}. \end{aligned} \quad (1)$$

Upon calculating all distances among rectangles, we select the minimum distances between clusters and accumulate those to represent *distance* Σ_d between two consecutive frames using the following algorithm.

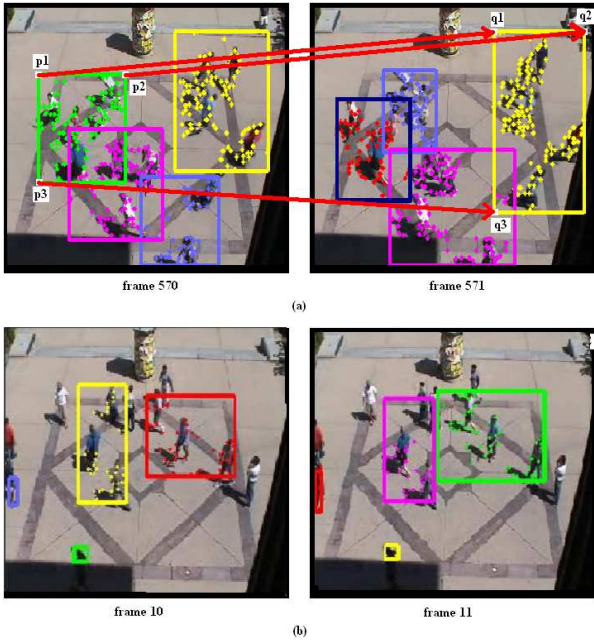


Figure 3: K-means clustering: (a) abnormal motion and (b) normal motion. Red arrows in (a) show Euclidean distances among points $p_1, p_2, p_3, q_1, q_2,$ and q_3 between rectangles p & q of two consecutive video frames.

Algorithm.

[F : total number of rectangles in any frame f , S : total number of rectangles in frame $f + 1$, m : rectangle counter in frame f , n : rectangle counter in frame $f + 1$]

1. *begin*
2. *Initialization:* $m = 1, n = 1$
3. *if* $m \leq F$
4. *then:* *if* $n \leq S$
 - then:*
 - begin*
 - (i) *using Eq. 1 calculate* Γ *between two rectangles and store*
 - (ii) *increment of* n *by 1*
 - (iii) *go to step 4*
 - end*
 - else:*
 - begin*
 - (i) *select minimum* Γ *and store*
 - (ii) *ignore corresponding rectangles*
 - (iii) *decrement of* F *by 1*
 - (iv) *increase* m *by 1 and set* $n = 1$
 - (v) *go to step 3*
 - end*
5. *else:* *sum up all the stored minimum distances* Γ *to get* Σ_d
6. *end*

As compare to normal motion, the positions and sizes of the rectangles of abnormal motion are noticeably differ between two consecutive frames (see figure 3 (a)). Consequently, the value of Γ and hence is the Σ_d will be higher. As compare to abnormal motion, the positions and sizes of rectangles of normal motion are almost similar between two consecutive frames (see figure 3 (b)). As a result, the value of Γ and hence is the Σ_d will be smaller.

3.2.3 Threshold estimation

A predefined threshold T_p value can differentiate each frame with respect to its assigned distance value whether its motion is normal or abnormal. There are several methods which may apply to estimate T_p . One of the approaches of computing T_p is that we consider the maximum number of distances in large videos that contain exclusively normal motions:

$$T_p = \max_{h=1..t} \{\Sigma_d\}_h \quad (2)$$

where t is the number of frames of the video database. Any frame having value of Σ_d which is greater than the T_p will be considered as abnormal motion frame. The T_d depends on the controlled environment, namely the distance of the camera to the scene, the orientation of the camera, the type and the position of the camera, lighting system, density of the crowd, etc. The more is the distance of the camera to the scene, the less is the quantity of optical flows and blobs. Taking into account of these facts, we consider that we have at least one threshold by a video stream. If we have M video streams, which are the case in sites such as play ground, concert, parking place, town center, political event, etc., then we select at least M thresholds. If the environment changes, then the threshold should be regenerated.

4. EXPERIMENTAL RESULTS

To conduct the experiment, as a data set we used different videos from different outdoor places which comprise of both normal and abnormal motions. The experimental results of one of the videos have been presented in figure 4. The video consists of 657 frames (attribute 320×240) where both normal and abnormal motion exist. Abnormal motion includes a sudden situation when a group of people start running. From frame 1 to 550 the motion of people is normal. People tends to run about frame number 551. More precisely, the assigned distance of frame 551 will be higher than any other before assigned distances among frames. Consequently, this frame can be considered as the ground truth frame. Ground truth is the process of manually marking what an algorithm is anticipated to output.

In figure 4 the blue clouded curve is the output of the proposed approach. The Gaussian like curve represents the abnormal motion when the group of people is trying to leave the place with very quick motion. To illustrate the performance of the algorithm, the ground truth frame and an arbitrary next video frame and their corresponding positions on the output curve have been indicated by green and red arrows respectively.

5. CONCLUSION

We proposed a method that estimates sudden changes of motion variations of a set of interest points. Optical flow information is computed from those points of interest. The

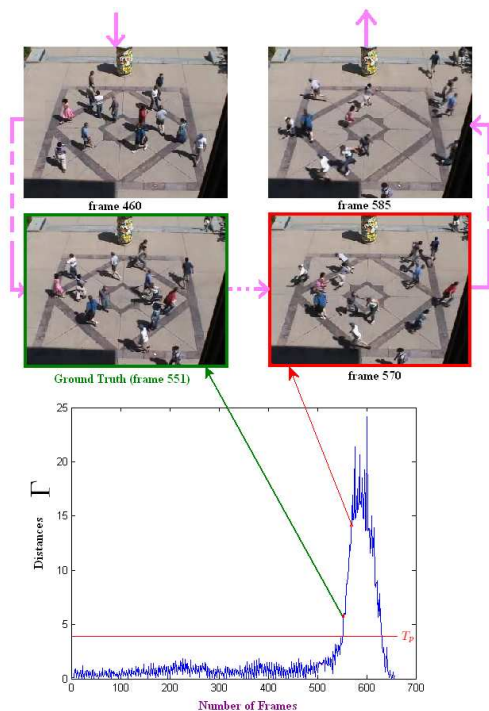


Figure 4: The blue curve depicts the output of the proposed approach and horizontal red line indicates threshold T_p . The curve above the line T_p represents abnormal motion.

k-means algorithm clusters the obtained optical flow information to get the distance between two consecutive frames. A predefined threshold T_p decides each frame whether the frame belongs to normal motion frame or abnormal motion frame. Although the performance of the approach is promising because of its robustness, it is necessary to define T_p for getting first rate results.

It is noticeable that the lighting condition which causes specially shadows of moving bodies has a severe effect on the subsection of detection of motion. In the proposed approach this effect did not take into account. But it would be worth interesting to consider the effect in many applications of computer vision. As a future work, this effect could be taken into account and minimized. Future work would also include to maintain the algorithm for particular applications in airports, subways, train stations, banks, cinema halls, schools, supermarkets, hospitals, and etc to detect abnormal and/or emergencies cases.

6. ACKNOWLEDGEMENTS

This work was supported by the MIAUCE project, which is the 6th Framework Research Programme of the European Union (IST-2005-5-033715). <http://www.miauce.org>.

7. REFERENCES

[1] E. Andrade, S. Blunsden, and R. Fisher. Hidden markov models for optical flow analysis in crowds. *ICPR'06. 18th International Conference*, vol. 1:460–463, 2006.

[2] 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.

[3] O. Boiman and M. Irani. Detecting irregularities in images and in video. *International Journal of Computer Vision*, 74(1):17–31, 2007.

[4] R. M. C. Schmid and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 2(37):151–172, June 2000.

[5] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–152, 1988.

[6] S. P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–136, 1982.

[7] 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.

[8] 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.

[9] 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.

[10] H. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. In *Technical Report CMU-RI-TR-3*. Carnegie-Mellon University, Robotics Institute, 1980.

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

[12] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.