Fast abnormal event detection from video surveillance

S. Javanbakhti, S. Zinger, P. H.N. de With
Electrical Engineering department, Eindhoven, The Netherlands

Abstract—Video surveillance systems are becoming increasingly important both in private and public environments to monitor activity. In this context, this paper presents a novel block-based approach to detect abnormal situations by analyzing the pixel-wise motion context, as an alternative for the conventional object-based approach. We proceed directly with event characterization at the pixel level, based on motion estimation techniques. Optical flow is used to extract information such as density and velocity of motion. The proposed approach identifies abnormal motion variations in regions of motion activity based on the entropy of Discrete Cosine Transform coefficients. We aim at a simple block-based approach to support a real-time implementation. We will report successful results on the detection of abnormal events in surveillance videos captured at an airport.

Keywords: Video surveillance systems, optical flow, Discrete Cosine Transform, event detection, real-time

1. Introduction

Recently, government agencies, businesses, and even schools are turning toward video surveillance as a means to increase public security. Video surveillance has been a key component in ensuring security at airports, banks, casinos, and correctional institutions [1]. The goal of visual surveillance is not only to use cameras instead of human eyes, but also to accomplish the entire surveillance task as automatically as possible using video analysis. Visual surveillance in dynamic scenes using video analysis has a wide range of potential applications, such as security for infrastructures and important buildings, traffic surveillance in cities and highways, detection of military targets, etc. [2]. For automatic dynamic scene analysis, anomaly detection is a challenging task especially given a scene consisting of complex correlated activities of multiple objects [3]. Anomaly detection techniques can be divided into two broad families of approaches, namely pattern-recognition-based and machine-learning-based methods. The pattern recognition approaches are typically those where the type of abnormal activity or object is a priori known. Nevertheless, such shape recognition methods require a list of objects or behavior patterns that are anomalous. Unfortunately, this is not always possible, especially where suspicious activities cannot be known in advance. An alternative approach is based on learning "normal" behavior from a video sequence exhibiting regular activity and then flag moving objects whose behavior deviates from normal behavior [4]. As discussed in different review papers [5] [6], many methods implement a general pipeline-based framework: at first moving objects are detected, then they are classified and tracked over a certain number of frames and finally, the resulting paths are used to distinguish "normal" behavior of objects from the "abnormal" [7] [8]. Although tracking-based methods have proven successful in different applications, they suffer from fundamental limitations. First, implementing such pipeline methods can result in a fragile architecture which may suffer from an error propagating through the subsequent processing stages. Secondly, tracking multiple objects at the same time requires complex algorithms and is computationally heavy. Therefore, multi-object tracking is not always efficient in crowded areas where objects regularly fully or partially occlude each other. This task is spatially hard in surveillance videos where quality and color information can be poor. Thirdly, tracking is efficient mostly with rigid moving bodies such as cars, trains, or pedestrians, and is not well suited to deal with unstructured motion such as waves on the water or trees shaking due to wind [4]. To address these limitations, some authors have recently proposed learning methods based on characteristics other than motion paths [9]. In such a case, there is no need for object tracking, instead, we consider pixel-level features. The main idea is to analyze the general motion context instead of tracking subjects one by one.

We design a general framework based on features directly extracted from motion such as velocity at pixel-level. This will lead to an image that expresses the motion in the scene. Then we analyze the information content of that image in the “frequency” domain by computing the entropy of the involved DCT coefficients that are related to the local motion description. After successfully analyzing motion in each frame, we should understand the behavior of the objects. Behavior understanding involves the analysis and recognition of motion patterns, and the description of actions and interactions at high level [2]. For an event to be considered normal or abnormal based on motion features, we compare the entropies for each block to the median averaged values over time to classify events into normal and abnormal. For the complete system, we aim at simple and fast methods suitable for real-time operation and allowing parallel processing due to block-based computing. The real-time aspect also explains why we have adapted the DCT, as this transform can be efficiently implemented.

The sequel of the paper is as follows. Section 2 describes the approach for abnormal event detection including motion...
estimation, measuring entropy and then detecting abnormal events. Section 3 presents our experimental results and the comparison with the state-of-the-art algorithm [9]. Section 4 concludes the paper.

2. Abnormal event detection algorithms

Our abnormal event detection is based on motion features extracted with a motion estimation technique. Motion estimation in image sequences aims at detecting regions corresponding to moving objects such as vehicles and humans. Detecting moving regions provides features for later processing such as tracking and behavior analysis. We base our features on pixel-based optical flow, since this is the most natural technique for capturing motion independent of appearance [10]. Computer vision experience suggests that computation of optical flow is not very accurate, particularly on coarse and noisy data. To deal with this, we use optical flow at each frame using the Lucas-Kanade algorithm [11]. Optical-flow-based motion estimation uses characteristics of flow vectors of moving objects over time to detect moving regions in an image sequence, relating each image to the next. Each vector represents the apparent displacement of each pixel from image to image [12]. The result of optical flow is the value of displacement of each pixel at both vertical and horizontal direction. We combine this displacement to obtain a motion magnitude vector. To process these motion vectors, we substitute pixel values for the estimated motion and we divide each frame into blocks. We expect that during abnormal events the motion patterns and therefore the energy of the images containing motion vectors change compared to normal behavior. We apply a DCT to each block, as the DCT provides a compact representation of the signal’s energy. Then we compute the entropy of the DCT coefficients to measure the information content of the DCT coefficients [13]. The entropy is defined as [14]:

\[ E = - \sum_{i=1}^{N} p \log p, \]

where \( N \) is the size of image and \( p \) contains the probability of the motion intensity value at a certain pixel location. The number of bins in the histogram is specified by the image type. In our case the motion magnitude per pixel forms a gray-scale image and we use 256 bins which correspond to the number of gray levels. For deciding whether the event is normal or abnormal, we compare the entropy value with thresholds which we learn per block in the beginning of the video sequence. This threshold is based on a median value of the entropies which we estimate during the first 500 frames of the video. Obviously, we assume that abnormalities will not occur during the first 500 frames of the surveillance sequence. In general this median computation of the threshold can be done continuously through the video because the abnormalities will be filtered out by taking a median value. We limit the median filtering over time for controlling the total complexity of the calculations. An abnormal event is indicated when the value of the entropy for the current frame is higher than the threshold defined for that block. An alternative approach [9] used for comparison defines a statistic measure that describes how much the optical flow vectors are organized or cluttered in the frame. That algorithm uses a metric which is the scalar product of the normalized values of the following factors: direction variance, motion magnitude variance and direction histogram peaks. This metric is then compared to a threshold which is a manually set at a constant value. Figure 1 illustrates our block-based processing framework in dynamic scenes.

![Block-based processing framework](image)

Fig. 1: Block-based processing framework.

3. Experimental results

In our experiments, we use surveillance videos of an airport where several abnormal situations are simulated by a group of volunteers. These situations include running of several people to the middle and from the middle of the scene, dancing in a circle and forming a fast growing group (overcrowding). In total, we have 6 of such a situation which occur in the video sequence of 12,000 frames. The spatial resolution of an original video frame is 720 × 576 with 29 frames per second, which is typical for surveillance in broadcast quality. We implement our block-based approach which is mentioned above on these data set. Furthermore, we compare the performance of our approach with an alternative approach presented by Ihaddadene and Djeraba [9] and we analyze the obtained results.

3.1 Performance of our approach

In our current implementation, we divide each frame into 4 blocks for simplicity but this can be easily changed. For each block we calculate the entropy of the DCT coefficients of the motion vector magnitudes and then compute the median value over the first 500 frames. Based on experiments and evaluation, the threshold for the median entropy to classify an abnormal event is empirically set to 3 times the median value. Abnormality indicator for the whole frame is raised if abnormality is detected in any of the blocks. Figure 2(a) shows an example of a frame of our videos and in Figure 2(b) we indicate the presence of the alarm. Figure 3 illustrates the result of our system in abnormal
situation. The original frame of the airport video sequence shows a situation where people are moving in a circle and the analysis window shows a clear warning with the 4 bars on all 4 blocks. From this video sequence, our approach detected 5 out of 6 abnormal events, without producing false alarms. Figure 4(a-d) also show frames of abnormal situations. Figure 5(a) visualizes ground-truth results in block number 3 and Figure 5(b) illustrates how our system works in those intervals in that block.

3.2 Result of comparison

We have applied the approach of Ihaddadene and Djeraba [9] to our video sequences. In airports, people normally move in any desired direction, therefore we do not consider their movement direction as a suitable feature which has been used in [9]. After experimenting with the original approach of these authors, we have tried to improve those results by considering only the motion variance per frame. We have experimented with several threshold values for this approach and the best results obtained are illustrated in Figure 5(c). We can observe that this method produces false alarms in normal situations and it has missed 3 abnormal events.

4. Conclusions

In this paper, we have developed a motion-context-based algorithm to detect abnormal events in surveillance videos of a public place. Our major contribution is introducing informative features based on motion and using an automatically updated threshold to detect abnormal events. We have discovered that the entropy of the DCT-transformed motion magnitude is a reliable measure for classifying whether the current activity in the video is normal or not. Because the proposed method is block-based, we can indicate exactly in which part of the frame the abnormal event takes place. Another advantage of using block is the possibility for parallel processing in a real-time implementation, since each block can be independently processed. The type of signal processing is chosen that fast real-time analysis is enabled. We obtain processing results on an airport surveillance video by detecting 5 out of 6 abnormal events while preventing false alarms. Our framework is generic and does not depend on the type of scene. For future work, we envisage more research on DCT-based features for abnormal event detection. We would also like to test our approach on different types of scenes where not only humans are involved but also moving objects.

References

Fig. 5: (a) Ground-truth results in block number 3 for the complete surveillance video, (b) warnings based on our detection corresponding to those events (bars show alarm in block number 3 at current frame and first two bars correspond to one event) and (c) warning for comparison corresponding to the system of [9].