Detection of Visual Abnormal Events via One-class SVM

Tian WANG¹, Hichem SNOUSSI¹, and Fethi SMACH²

¹Institut Charles Delaunay-LM2S-UMR STMR 6279 CNRS, Université de Technologie de Troyes, France
²ActivNetworks, 1 Rue Terre Neuve 91944 Courtabœuf, France

Abstract—In this paper, we propose an algorithm to detect abnormal events based on video streams. The algorithm is based on optical flow descriptor and one-class SVM classifier. The optical flow is computed at each pixel of the video frame, and the one-class SVM, after a learning period characterizing normal behavior, detects the abnormal blobs in the current frame. As the algorithm is not based on object tracking, it can work in crowded scenes where the tracking-based methods might fail. Extensive testing on benchmark datasets corroborates the effectiveness of the proposed detection method.

Keywords: abnormal detection; optical flow; one-class SVM

1. Introduction

One of the most popular tasks in computer vision is analyzing abnormal events in crowded scenes. The abnormal detection methods can be divided into two categories: tracking-based methods and feature-based methods. The methods based on tracking might be influenced by the scene environment. As the amount of moving objects in the scenes is large, or features of the target are confused by other objects, the accuracy of the algorithm will be reduced. Thus the feature-based methods are more robust and work better in crowded scenes.

Approaches which use behavioral model without the need of tracking procedure were described in [1], [2], [3], [4]. The detected foreground blobs are analyzed as scene events and then behavior patterns are represented by Bayesian models. Abnormal events are detected based on the statistical feature of observed behavior pattern. Successful results were obtained on several scenes, but the training which depends on large amounts of normal behavior patterns is the bottleneck.

Some authors focus on spatio-temporal models. Markov random field framework was used in [5] to describe co-occurrence matrix which is built based on 3D spatio-temporal foreground mask feature. In [6], multiple temporal features are extracted from foreground blobs and then confined in SVM based models for abnormal detection. In [7], probabilistic Latent Semantic Analysis (pLSA) is applied in the field of visual features to detect abnormal activity patterns.

Biological models have been introduced for abnormal detection. Neural network is used in [8] to model the superior colliculus (SC) to detect anomalies in panoramic image.

Low level motion features could also be used for abnormal event detection. In [9], a semi-unsupervised learning method based on optical flow is introduced and latent Dirichlet allocation model is used to detect abnormal moving objects. An algorithm based on collecting low-level statistics is presented in [10]. A method to detect local abnormal events and global abnormal events via sparse reconstruction is introduced in [11].

This paper presents a feature-based abnormal detection method. The proposed algorithm is composed of two parts, low-level visual features extraction, and events classification by one-class SVM. By avoiding tracking objects, the proposed method can work in a robust way in crowded scenes. Moreover, the SVM method shows high performance results only based on normal behavior frames. The rest of the paper is organized as follows. In Section 2, motion features are introduced. In Section 3, the one-class SVM classification method is presented. In Section 4, we describe an overview of our visual-based abnormal detection method. In Section 5, we present experiment results. Finally, conclusions are presented in Section 6.

2. Features selection

Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image. It can give important information about the spatial arrangement of the objects and the change rate of this arrangement [12]. Abnormal action can be exhibited by the direction and the amplitude of the movement, we chose optical flow as description finding features of the scene. B.Horn and B. Schunck [12] proposed the algorithm introducing a global constraint of smoothness to computer optical flow. The basic Horn-Schunck (HS) optical method is used in our work. The HS method combines a data term that assumes constancy of some image property with a spatial term that models how the flow is expected to vary across the image [13]. For two-dimensional image sequences, the optical flow is formulated as a global energy functional:

$$E = \int \int [(I_xu + I_yv + I_t)^2 + \alpha (\|\nabla u\|^2 + \|\nabla v\|^2)] dx dy$$ (1)

where \(I_x, I_y\) and \(I_t\) are the derivatives of the image intensity values along the \(x, y\) and time \(t\) dimensions respectively, \(u, v\) are the components of the optical flow, and \(\alpha\) is a
regularization constant. In order to minimize the functional E. Lagrange equations are used:

\[
\begin{cases}
I_u(I_u u + I_y v + I_t) - \alpha^2 \Delta u = 0 \\
I_y(I_u u + I_y v + I_t) - \alpha^2 \Delta v = 0
\end{cases}
\]

Subject to,

\[
\begin{cases}
\Delta u(x, y) = \nabla(x, y) - u(x, y) \\
\Delta v(x, y) = \nabla(x, y) - v(x, y)
\end{cases}
\]

where \(\nabla, \nabla\) are weighted average of \(u, v\) calculated in a neighborhood around the pixel location. As the solution depends on the neighboring values, when the neighbor pixels are updated, the solution need to be iterated. In an iterative scheme, the optical flow can be written as:

\[
\begin{cases}
u^{k+1} = \nabla^k - \frac{I_x(I_u \nabla + I_y \nabla + I_t)}{\alpha^2 + \beta + I_t^2} \\
v^{k+1} = \nabla^k - \frac{I_y(I_u \nabla + I_y \nabla + I_t)}{\alpha^2 + \beta + I_t^2}
\end{cases}
\]

where \(k\) is the last calculated result. The optical flows at normal and abnormal frames are shown in Fig. 1. In this example a single time step was taken for normal scene and abnormal scene, so that the computations are based on just two adjacent images.

3. Classification

Support Vector Machine (SVM) is a method based on statistical learning theory and risk minimization for classification and regression. SVM is initially proposed by Vapnik and Lerner [14]. Later, SVM has been extended to nonlinear framework with the introduction of kernel methods [15], [16], [17]. The theory behind SVM is briefly presented below.

SVM theory is developed from the application of statistical learning theory results to linear classifiers. Suppose the case of two-class hyperplane classifiers in space \(H\):

\[
\langle w, x \rangle + b = 0 \quad w, x \in H, \quad b \in \mathbb{R}
\]

corresponding to the decision function:

\[
f(x) = \text{sgn}(\langle w, x \rangle + b)
\]

The hyperplane is depicted in Fig. 2. Statistical learning theory states that the optimal classifier can be found by maximizing the margin [16]. This can be expressed as a minimization problem:

\[
\min_w \frac{1}{2} \|w\|^2
\]

Subject to,

\[
y_i(\langle w, x \rangle + b) \geq 1, \quad i = 1, \ldots, m
\]

where \(m\) is the number of training data and \(y_i \in \{-1, +1\}\) is the label for the elements.

In one-class classification problem, it is assumed that the data from only one class, the positive class, are available. The one-class SVM frameworks suit the specificity of the abnormal event detection where one can just obtain samples in the normal scenes. In one-class SVM, the objective is to find out an appropriate region in the space \(X\) which contains most of the data drawn from an unknown probability distribution \(P\). This can be obtained by searching for a decision hyperplane in the feature space \(H\), which maximizes its distance from the origin, while only a small fraction of data falls between the hyperplane and the origin [16]. The problem can be presented as a constrained minimization one:

\[
\min_{w, \xi, b} \frac{1}{2} \|w\|^2 + \frac{1}{\nu} \sum_{i=1}^n \xi_i - b
\]

Subject to,

\[
\langle w, \Phi(x_i) \rangle \geq b - \xi_i, \quad \xi_i \geq 0
\]

where \(x_i \in X, i \in [1 \ldots n]\) are \(n\) training data in the data space \(X\). \(\Phi : X \to H\) maps vectors \(x_i\) in the feature space \(H\). \(\langle w, \Phi(x_i) \rangle - b = 0\) is the decision hyperplane, \(\xi_i\) is the slack variable for penalizing the outliers. \(\nu \in (0, 1]\) is the weight for controlled slack variable, it tunes the number
of acceptable outliers. $\Phi$ is defined for solving the non-linear classification problems. It’s a map from the nonempty set of the original input data $X$ to a feature space $H$ where the classification problem has a linear solution. For computing dot products in $H$, the kernel function is defined as $k(x, x') = \Phi(x) \cdot \Phi(x')$. Introducing the Lagrangian multipliers $\alpha_i$, the minimization problem can be changed to its dual form:

$$\min_{\alpha} \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j k(x_i, x_j)$$

subject to,

$$0 \leq \alpha_i \leq \frac{1}{m} \quad \sum_{i=1}^{n} \alpha_i = 1$$

Lagrange multiplier $\alpha_i$ can be found by solving the problem with standard quadratic programming methods, and $w$ is given by $w = \sum_{i=1}^{n} \alpha_i \Phi(x_i)$. When $\alpha_i \neq 0$, for vector $\Phi(x_i)$, the equations holds:

$$b - \xi_i = \langle w, \Phi(x_i) \rangle = \sum_{j=1}^{n} \alpha_j k(x_i, x_j)$$

The decision function in the data space $X$ is defined as:

$$f(x) = \text{sgn}(\sum_{i=1}^{n} \alpha_i k(x, x_i) - b)$$

B. Schölkopf and A. Smola [18] have proven that if proper parameters are given, the traditionally used kernels, such as Gaussian, polynomial, and sigmoidal kernel, have similar performances. As we deal with spatial features, Gaussian kernel seems a practicable choice, it’s a semi-positive definite kernel which satisfies the condition of Mercer [19], [20].

$$k(x_i, x_j) = \exp(-\frac{||x_i - x_j||^2}{2\sigma^2}), (x_i, x_j) \in X \times X$$

The parameter $\sigma$ indicates the scale factor at which the data should be clustered.

4. The proposed method

In this section, we describe a method for detecting abnormal events in video streams. Assume that a set of frames $[I_1 \ldots I_n]$ in which the person is walking or loitering, are considered as normal events, as shown in Fig. 1(a). The frames in which the person is running or walking with a sudden split are regarded as abnormal events, as shown in Fig. 1(b). The objective is to detect the abnormal events based on training samples via one-class SVM. The general architecture of the abnormal detection method is presented in Fig. 3, and outlined in the following.

Step 1: The original incoming frames are processed via Horn-Schunck (HS) optical flow method to get the features for classification at every pixel. The optical flow features are computed at gray scale.

**Step 2:** One-class SVM is used to classify feature samples of incoming video frames. Three strategies are chosen for obtaining the features of the image. The sketch image for choosing the features is shown in Fig. 4.

Method 1: Take the optical flow at each pixel of the image as feature samples, as shown in Fig. 4(a). The normal and abnormal events are people action. In the dataset UMN [21], normal event is walking or loitering, and abnormal event is running. The video sequence in our work is labeled as frames in which the person is walking or loitering, are considered as normal events, as shown in Fig. 1(a). The frames in which the person is running or walking with a sudden split are regarded as abnormal events, as shown in Fig. 1(b). The objective is to detect the abnormal events based on training samples via one-class SVM. The general architecture of the abnormal detection method is presented in Fig. 3, and outlined in the following.

**Method 2:** One-class SVM is used to classify feature samples of incoming video frames. Three strategies are chosen for obtaining the features of the image. The sketch image for choosing the features is shown in Fig. 4.

**Method 3:** Take the optical flow at each pixel of the image as feature samples, as shown in Fig. 4(a). The normal and abnormal events are people action. In the dataset UMN [21], normal event is walking or loitering, abnormal event is running. The video sequence in our work is labeled as frames in which the person is walking or loitering, are considered as normal events, as shown in Fig. 1(a). The frames in which the person is running or walking with a sudden split are regarded as abnormal events, as shown in Fig. 1(b). The objective is to detect the abnormal events based on training samples via one-class SVM. The general architecture of the abnormal detection method is presented in Fig. 3, and outlined in the following.

**Step 3:** Compute optical flow for each frame. The optical flow features are computed at gray scale.

**Step 4:** Detect abnormal events on whole image.
The image is also segmented into several blocks, as shown in Fig. 4(b), the image is separated into $p \times q$ blocks, $p$ is the number of blocks at the width dimension of the image, $q$ is the number of blocks at the length dimension of the image. The width of the block is $h$ pixels, the length of the block is $w$ pixels, there are $h \times w$ points in the block. The feature of block at $i_{th}$ row and $j_{th}$ column in the $k_{th}$ frame is noted as $F_{i,j,k}$. For each block, the feature $F$ is arranged by the optical flow of all the points in the form $[O_{1,1}, O_{2,1}, \cdots, O_{h,w}]$. For the video streams, take the features of block at normal images as the training samples for one-class SVM, and then detected block-by-block.

**Method 3**: The image is also segmented into several blocks, but the training samples are all the blocks at one image, as shown in Fig. 4(c). Similar as the Method 2, we separate one frame to $p \times q$ blocks, the size of each block is $h \times w$. At $k_{th}$ frame, the feature sample of all the blocks on this frame is $[block_{1,1,k}, block_{1,2,k}, \cdots, block_{p,q,k}]$, a vector of dimension $(p \times q) \times (h \times w)$. To get the training data at the normal sequence from $1_{th}$ to $m_{th}$, the data are arranged as $[block_{1,1,1}, block_{1,2,1}, \cdots, block_{p,q,1}, \cdots, block_{1,1,n}, \cdots, block_{p,q,n}]$, a vector of dimension $(p \times q \times n) \times (h \times w)$. For detecting, the test sample is the feature of one block. For example, the block at position $(i,j)$ at $m_{th}$ frame is noted as $block_{i,j,m}$.

The sequence which just has one person is taken as an example for detailing the algorithm performance. The scene is presented in Fig. 5. Four pictures in Fig. 5 show the scene without people, the person walking and the person running at different directions. For abnormal detection, the training sequence for studying by SVM in which the person is walking, such as Fig. 5(b) shows. The detected sequence in which the person is running, such as Fig. 5(c)(d) shows. The results of these three strategies are shown in Fig. 6. At Fig. 6(b)(c), the abnormal detections on the background are marked by white circles, they are taken as false alarms. The detected result of pixel-by-pixel feature selection strategy has the most false alarms. For pixel-by-pixel strategy just take the feature at one pixel, and is more susceptible to the optical flow changing. The feature chosen by block can get better detected result than pixel-by-pixel result. The block-by-block strategy which is shown in Fig. 6(c) take one block as the local monitor, it considers the situation of several pixels as an integrality. The block-by-block strategy is more robust than just taking one pixel as feature sample. Taking all the blocks on the image as the training samples has no false alarms and has similar detected results on the person, and it is the best way for abnormal detection of this sequence.

**Step 3**: As the abnormal detection problem is for analyzing the human action, the SVM detection result can be combined with foreground detection. The abnormal detections which are not on the foreground can be deleted, they are considered as noise of the detected results. The background subtraction method presented by O. Tuzel et al.[22] is adopted. In the method, each pixel is represented with layers of Gaussian distributions, and then recursive Bayesian learning is used to estimate the probability distribution of mean and covariance of each Gaussian[22]. This method can remove the shadows of moving persons. Then, a logical conjunction of the results of one-class SVM for optical flow and the foreground of the image is made. When the points or blocks are detected as abnormality and also from the foreground, they are considered as the abnormalities by the
global algorithm.

**Step 4:** After acquiring detected result of each point or each block, then the decision of global frame abnormality is made by presetting a number as threshold. If the number of abnormal points or blocks is larger than the threshold, we consider the frame as an abnormal one.

**Case 1:** If there are no abnormal detected points or blocks on the frame, this frame is considered as a normal one.

**Case 2:** If the number of abnormal points or blocks at the frame exceeds the threshold but this frame is labeled as a normal one, the detected result of the whole image via one-class SVM is considered as a false alarm.

**Case 3:** If the number of abnormal points or blocks on the frame exceeds the threshold and this frame is labeled as an abnormal one, then the detected result via one-class SVM is considered as correct detection.

## 5. Result

This section presents the results of experiments conducted to analyze the performance of the proposed method of detecting abnormal events. The normal and abnormal scenes are shown in Fig. 7.

The detected results of the lawn scene are shown in Fig. 8. The points marked with white color are the abnormal detections via SVM, the points marked with cyan color are the abnormal detections and also on the foreground. At Fig. 8 (b)(c), the abnormal detected results on the background is marked by white circles. Fig. 8(d) is the result taken all blocks at the whole image as the training samples, it has the best detected results.

We present one special situation of the events on lawn scene. As presented on Fig. 9, when most persons are running, on the lower part of the image, one person is walking. The people who is walking is cut out from the walking sequence at UMN dataset. The detected results of this special situation are shown in Fig. 9. The pixel-by-pixel and block-by-block feature selection strategies detect the walking person as abnormality. These two strategies just analyze information at the fixed pixel or block. At the lower part of image, there are no people on the training sequence, so the walking people is regarded as an abnormal event. The right strategy should be chosen depending on the application. If the region is ‘no admittance’, the walking person in this region is abnormal. The feature selected strategy can be pixel-by-pixel or block-by-block. If only the running movement is abnormal, the opportune strategy for feature selection should take all the blocks on the whole image as training samples. Fig. 9(d) has the less abnormal detections. Because the feature selection strategy which takes all the blocks on the image as training samples considers an overall situation, it has the largest robustness and the least sensitiveness. Moreover, at Fig. 9(b)(c)(d), the abnormal detected results are not on all the persons. As the frames are the beginning of the running events, the optical flow is not much different from walking. Some parts of these persons are detected as normal.

Other two tested scenes are an indoor sequence and a plaza sequence. The abnormal detected results of these two scenes are shown in Fig. 10. The detected results show that the pixel-by-pixel feature selection strategy is the most sensitive method for detecting. When taking the blocks at the whole image as the training samples, the detection is the most robust.

Performance summary on the UMN dataset compares with paper [6] is in TABLE 1. For these three scenes, we get approximative detection rate with paper [6], and the false alarms are ameliorated.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>lawn</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>plaza</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Table 1: Detection results of different scenes. DR='detection rate'; FPR='false positive rate'. The last two columns are the statistic results of the proposed method.

## 6. Conclusions

A method was proposed for abnormal detection. The method is based on two components, extracting optical flow of the points on the frame as the features, and applying one-class SVM for classification. Three features selection strategy are presented, the features samples are chosen pixel-by-pixel, block-by-block, and taking the blocks on the whole image as the training samples. It is a feature based approach without any tracking algorithm, so it might be more reliable in a crowded video than tracking based methods. The experimental results have proven the validity of the proposed approach.

### References

Fig. 7: three scenes of the test video sequences: (a)(b)(c) the first row is scene on the lawn; (d)(e)(f) the second row is scene indoor; (g)(h)(i) the third row is scene on a plaza; (a)(b)(d)(e)(g)(h) normal events, all the persons are walking; (c)(f)(i) abnormal events, all the persons are running.

Fig. 8: detected results of lawn scene: (a) the original image; (b) the abnormal detection by pixel-by-pixel; (c) the abnormal detection by block-by-block; (d) the abnormal detection by taken block on the whole image as training samples; (e) the dilative foreground of the image; (f) the abnormal detection by pixel-by-pixel and also on the foreground; (g) the abnormal detection by block-by-block and also on the foreground; (h) the abnormal detection by taken block on the whole image as training samples and also on the foreground.
Fig. 9: detected results of special situation in the lawn scene: (a) the original image of one person walks on the lower part of the image; (b) the abnormal detection by pixel-by-pixel strategy; (c) the abnormal detection by block-by-block strategy; (d) the abnormal detection by taken all the blocks on the whole image as training sample.


