

Research Article

Online Detection of Abnormal Events in Video Streams

Tian Wang,¹ Jie Chen,² and Hichem Snoussi¹

¹ *Institut Charles Delaunay, LM2S-UMR STMR 6279 CNRS, University of Technology of Troyes, 10004 Troyes, France*

² *Observatoire de la Côte d'Azur, UMR 7293 CNRS, University of Nice Sophia-Antipolis, 06108 Nice, France*

Correspondence should be addressed to Tian Wang; wangtian8704@gmail.com

Received 19 September 2013; Accepted 12 November 2013

Academic Editor: Yi Zhou

Copyright © 2013 Tian Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose an algorithm to handle the problem of detecting abnormal events, which is a challenging but important subject in video surveillance. The algorithm consists of an image descriptor and online nonlinear classification method. We introduce the covariance matrix of the optical flow and image intensity as a descriptor encoding moving information. The nonlinear online support vector machine (SVM) firstly learns a limited set of the training frames to provide a basic reference model then updates the model and detects abnormal events in the current frame. We finally apply the method to detect abnormal events on a benchmark video surveillance dataset to demonstrate the effectiveness of the proposed technique.

1. Introduction

Visual surveillance is one of the major research areas in computer vision. In a crowd image analysis problem, the scientific challenge includes abnormal events detection. For instance, Figure 1(a) illustrates a normal scene where the people are walking. In Figure 1(b), all the people are suddenly running in different directions. This dataset imitates panic-driven scenes.

Trajectory analysis of objects was described in [1–3]. The moving object was labeled by a blob in consecutive frames, and then a trajectory was produced. The deviation from the learnt trajectories was defined as abnormal events. Tracking based approaches are suitable for the sparse scenes with a few objects. The target might be lost due to occlusion.

In [4, 5], abnormal detection approaches which used features encoding motion, texture, and size of the objects were introduced. Local image regions in a video were analyzed by employing background subtraction method; then a dynamic Bayesian network (DBN) was constructed to model normal and abnormal behavior, and finally a likelihood ratio test was applied to detect abnormal behaviors. In [6], a space-time Markov random field (MRF) model which detected abnormal activities in a video was proposed, mixture of probabilistic principal component analyzers (MPPCA) was adopted to model local optical flow. The prediction is based

on probabilistic assumption techniques where an accurate model exists, but there are various situations where a robust and tractable model cannot be obtained; model-free methods are needed to be studied.

Spatiotemporal motion features described by the context of bag of video words were adopted to detect abnormal events. In [7], the authors presented an algorithm which monitored optical flow in a set of fixed spatial positions, and constructed a histogram of optical flow. The likelihood of the behavior in a new coming frame concerning the probability distribution of the statistically learning behavior was computed. If the likelihood fell below a preset threshold, the behavior was considered as abnormal. In [8], irregular behavior of images or videos was detected by an inference process in a probabilistic graphical model. In [9, 10], the video pixels were densely sampled to form the feature. These methods are based on the partial information of images, such as small blocks in a frame, without fully exploiting the global information of the feature. In [11–13], spatiotemporal features modeled motion regions of the frame as background, and anomaly was detected by subtracting the newly sample to the background template. These works are similar to the change detection method when the background is not stable.

In this paper, the proposed algorithm is composed of two parts. Firstly, a covariance feature descriptor is constructed over the whole video frame, and then a nonlinear one-class

support vector machine algorithm is applied in an online fashion in order to detect abnormal events. The features are extracted based on the optical flow which presents the movement information. Experiments of real surveillance video dataset show that our online abnormal detection techniques can obtain satisfactory performance. The rest of the paper is organized as follows. In Section 2, covariance matrix descriptor of motion feature is introduced. In Section 3, the online one-class SVM classification method is presented. In Section 4, two abnormal detection strategies based on online nonlinear one-class SVM are proposed. In Section 5, we present results of real-world video scenes. Finally, Section 6 concludes the paper.

2. Covariance Descriptor of Frame Behavior

The optical flow is a feature which presents the direction and the amplitude of a movement. It can provide important information about the spatial arrangement of the objects and the change rate of this arrangement [14]. We adopt Horn-Schunck (HS) optical flow computation method in our work. The optical flow of the gray scale image is formulated as the minimizer of the following global energy functional:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2) \right] dx dy, \quad (1)$$

where I is the intensity of the image, I_x , I_y , and I_t are the derivatives of the image intensity value along the x , y , and time t dimension, u and v are the components of the optical flow in the horizontal and vertical direction, and α represents the weight of the regularization term.

We introduce the covariance matrix encoding the optical flow and intensity of each frame as the descriptor to represent the movement. The covariance feature descriptor is originally proposed by Tuzel et al. [15] for pattern matching in a target tracking problem. The descriptor is defined as

$$F(x, y, i) = \phi_i(I, x, y), \quad (2)$$

where I is the color information of an image (which can be gray, RGB, HSV, HLS, etc.), ϕ_i is a mapping relating the image with the i th feature from the image, F is the $W \times H \times d$ dimensional feature extracted from image I , W and H are the image width and image height, and d is the number of chosen features. For each frame, the feature can be represented as $d \times d$ covariance matrix:

$$\mathbf{C} = \frac{1}{n-1} \sum_{k=1}^n (\mathbf{z}_k - \boldsymbol{\mu})(\mathbf{z}_k - \boldsymbol{\mu})^\top, \quad (3)$$

where n is the number of the pixels sampled in the frame, \mathbf{z}_k is the feature vector of pixel k , $\boldsymbol{\mu}$ is the mean of all the selected points, and \mathbf{C} is the covariance matrix of the feature vector F . The covariance descriptor \mathbf{C} of each frame dose not have any information regarding the sample ordering and the number of points [15]. Because the feature F can be designed as different approaches to fuse features, the covariance matrix descriptor proposes a way to merge multiple parameters. Different choices of feature vectors extraction are shown in

TABLE 1: Different choices of feature F to construct the covariance descriptor.

Feature vector F	
$F_1 (6 \times 6)$	$[y \ x \ u \ v \ u_x \ u_y]$
$F_2 (6 \times 6)$	$[y \ x \ u \ v \ v_x \ v_y]$
$F_3 (8 \times 8)$	$[y \ x \ u \ v \ u_x \ u_y \ v_x \ v_y]$
$F_4 (12 \times 12)$	$[y \ x \ u \ v \ u_x \ u_y \ v_x \ v_y \ u_{xx} \ u_{yy} \ v_{xx} \ v_{yy}]$
$F_5 (17 \times 17)$	$[y \ x \ u \ v \ u_x \ u_y \ v_x \ v_y \ u_{xx} \ u_{yy} \ v_{xx} \ v_{yy} \ I \ I_x \ I_y \ I_{xx} \ I_{yy}]$

Table 1, where I is the intensity of the gray image, u and v are horizontal and vertical components of optical flow, I_x , u_x , and v_x are the first derivatives of the intensity, horizontal optical flow, and vertical optical flow in the x direction respectively, I_y , u_y , and v_y are the first derivatives of the corresponding feature in the y direction and respectively, I_{xx} , u_{xx} , and v_{xx} are the second derivatives in x direction, I_{yy} , u_{yy} , and v_{yy} are the second derivatives in y direction. The flowchart of covariance matrix descriptor computation is shown in Figure 2. The optical flow and corresponding partial derivative characterize the interframe information or can be regarded as the movement information. The intensity of the frame and partial derivative of the intensity describe the intraframe information; they encode the appearance information of the frame.

If proper parameters are given, the traditionally used kernel, such as Gaussian, polynomial, and sigmoidal kernel, has similar performances [19]. Gaussian kernel $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ is chosen for our spatial features. The covariance matrix is an element in Lie group; the Gaussian kernel on the Euclidean spaces is not suitable for the covariance descriptors. The Gaussian kernel in Lie Group is defined as [20, 21]:

$$\kappa(\mathbf{X}_i, \mathbf{X}_j) = \exp\left(-\frac{\|\log(\mathbf{X}_i^{-1}\mathbf{X}_j)\|}{2\sigma^2}\right), \quad (\mathbf{X}_i, \mathbf{X}_j) \in G \times G, \quad (4)$$

where \mathbf{X}_i and \mathbf{X}_j are matrices in Lie Group G .

3. Online One-Class SVM

The essence of an abnormal detection problem is that only normal scene samples are available. The one-class SVM framework is well suitable to an abnormal detection problem. Support vector machine (SVM) is initially proposed by Vapnik and Lerner [22, 23]. It is a method based on statistical learning theory and has fine performance to classify data and recognize patterns. There are two frameworks of one-class SVM, one is support vector data description (SVDD) which is presented in [24, 25] and the other is ν -support vector classifier (ν -SVC) introduced in [26]. The SVDD formulation is adopted in our work. It computes a sphere shaped decision boundary with minimal volume around a set of objects. The



FIGURE 1: Examples of the normal and abnormal scenes: (a) Normal lawn scene: all the people are walking. (b) Abnormal lawn scene, all the people are running.

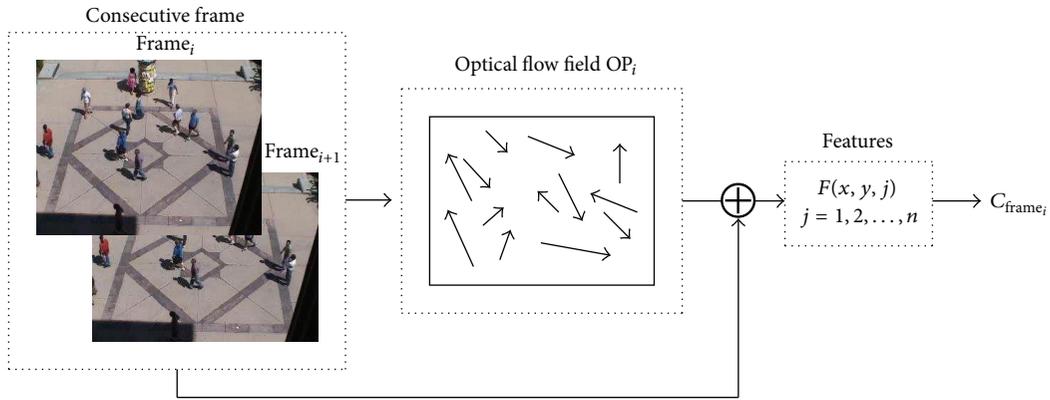


FIGURE 2: Covariance descriptor computation of a video frame.

center of the sphere \mathbf{c} and the radius R are to be determined via the following optimization problem:

$$\min_{R, \xi, \mathbf{c}} R^2 + C \sum_{i=1}^n \xi_i, \quad (5)$$

$$\text{subject to } \|\Phi(\mathbf{x}_i) - \mathbf{c}\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \forall i, \quad (6)$$

where n is the number of training samples and ξ_i is the slack variable for penalizing the outliers. The hyperparameter C is the weight for restraining slack variables; it tunes the number of acceptable outliers. The nonlinear function $\Phi: \mathcal{X} \rightarrow \mathcal{H}$ maps a datum \mathbf{x}_i into the feature space \mathcal{H} ; it allows to solve a nonlinear classification problem by designing a linear classifier in the feature space \mathcal{H} . κ is the kernel function for computing dot products in \mathcal{H} , $\kappa(\mathbf{x}, \mathbf{x}') = \langle \Phi(\mathbf{x}), \Phi(\mathbf{x}') \rangle$. By introducing Lagrange multipliers, the dual

problem associated with (6) is written by the following quadratic optimization problem:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i,j=1}^n \alpha_i \alpha_j \kappa(\mathbf{x}_i, \mathbf{x}_j), \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i = 1, \\ & \mathbf{c} = \sum_{i=1}^n \alpha_i \Phi(\mathbf{x}_i). \end{aligned} \quad (7)$$

The decision function is

$$\begin{aligned} f(\mathbf{x}) = \text{sgn} \left(R^2 - \sum_{i,j=1}^n \alpha_i \alpha_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \right. \\ \left. + 2 \sum_{i=1}^n \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}) - \kappa(\mathbf{x}, \mathbf{x}) \right). \end{aligned} \quad (8)$$

For the large training data, the solution cannot be obtained easily, and an online strategy to train the data is

used in our work. Let $\mathbf{c}_{\mathfrak{D}}$ denotes a sparse model of the center $\mathbf{c}_n = (1/n) \sum_{i=1}^n \Phi(\mathbf{x}_i)$ by using a small subset of available samples which is called dictionary:

$$\mathbf{c}_{\mathfrak{D}} = \sum_{i \in \mathfrak{D}} \alpha_i \Phi(\mathbf{x}_i), \quad (9)$$

where $\mathfrak{D} \subset \{1, 2, \dots, n\}$, and let $N_{\mathfrak{D}}$ denote the cardinality of this subset \mathfrak{D} . The distance of a mapped datum $\Phi(\mathbf{x})$ with respect to the center $\mathbf{c}_{\mathfrak{D}}$ can be calculated by

$$\begin{aligned} \|\Phi(\mathbf{x}) - \mathbf{c}_{\mathfrak{D}}\| &= \sum_{i, j \in \mathfrak{D}} \alpha_i \alpha_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \\ &\quad - 2 \sum_{i \in \mathfrak{D}} \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}) + \kappa(\mathbf{x}, \mathbf{x}). \end{aligned} \quad (10)$$

A modification of the original formulation of the one-class classification algorithm that consists of minimizing the approximation error $\|\mathbf{c}_n - \mathbf{c}_{\mathfrak{D}}\|$ is [27, 28]

$$\boldsymbol{\alpha} = \arg \min_{\alpha_i, i \in \mathfrak{D}} \left\| \frac{1}{n} \sum_{i=1}^n \Phi(\mathbf{x}_i) - \sum_{i \in \mathfrak{D}} \alpha_i \Phi(\mathbf{x}_i) \right\|^2. \quad (11)$$

The final solution is given by

$$\boldsymbol{\alpha} = \mathbf{K}^{-1} \boldsymbol{\kappa}, \quad (12)$$

where \mathbf{K} is the Gram matrix with (i, j) th entry $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ and $\boldsymbol{\kappa}$ is the column vector with entries $(1/n) \sum_{i=1}^n \kappa(\mathbf{x}_k, \mathbf{x}_i)$, $k \in \mathfrak{D}$.

In the online scheme, at each time step there is a new sample. Let $\boldsymbol{\alpha}_n$ denote the coefficients, \mathbf{K}_n denote the Gram matrix, and $\boldsymbol{\kappa}_n$ denote the vector, at time step n . A criterion is used to determine whether the new sample can be included into the dictionary. A threshold μ_0 is preset, for the datum \mathbf{x}_t at time step t , the coherence-based sparsification criterion [29, 30] is

$$\epsilon_t = \max_{i \in \mathfrak{D}} |\kappa(\mathbf{x}_i, \mathbf{x}_t)|. \quad (13)$$

First Case ($\epsilon_t > \mu_0$). In this case, the new data $\Phi(\mathbf{x}_{n+1})$ is not included into the dictionary. The Gram matrix $\mathbf{K}_{n+1} = \mathbf{K}_n$. $\boldsymbol{\kappa}_n$ changes online:

$$\begin{aligned} \boldsymbol{\kappa}_{n+1} &= \frac{1}{n+1} (n\boldsymbol{\kappa}_n + \mathbf{b}), \\ \boldsymbol{\alpha}_{n+1} &= \mathbf{K}_{n+1}^{-1} \boldsymbol{\kappa}_{n+1} \\ &= \frac{n}{n+1} \boldsymbol{\alpha}_n + \frac{1}{n+1} \mathbf{K}_n^{-1} \mathbf{b}, \end{aligned} \quad (14)$$

where \mathbf{b} is the column vector with entries $\kappa(\mathbf{x}_i, \mathbf{x}_{n+1})$, $i \in \mathfrak{D}$.

Second Case ($\epsilon_t \leq \mu_0$). In this case, the new data $\Phi(\mathbf{x}_{n+1})$ is included into the dictionary \mathfrak{D} . The Gram matrix \mathbf{K} changes:

$$\mathbf{K}_{n+1} = \begin{bmatrix} \mathbf{K}_n & \mathbf{b} \\ \mathbf{b}^\top & \kappa(\mathbf{x}_{n+1}, \mathbf{x}_{n+1}) \end{bmatrix}. \quad (15)$$

By using Woodbury matrix identity

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}, \quad (16)$$

\mathbf{K}_{n+1}^{-1} can be calculated iteratively:

$$\begin{aligned} \mathbf{K}_{n+1}^{-1} &= \begin{bmatrix} \mathbf{K}_n^{-1} & \mathbf{0} \\ \mathbf{0}^\top & 0 \end{bmatrix} + \frac{1}{\kappa(\mathbf{x}_{n+1}, \mathbf{x}_{n+1}) - \mathbf{b}^\top \mathbf{K}_n^{-1} \mathbf{b}} \\ &\quad \times \begin{bmatrix} -\mathbf{K}_n^{-1} \mathbf{b} \\ 1 \end{bmatrix} [-\mathbf{b}^\top \mathbf{K}_n^{-1} \mathbf{1}]. \end{aligned} \quad (17)$$

The vector $\boldsymbol{\kappa}_{n+1}$ is updated from $\boldsymbol{\kappa}_n$,

$$\boldsymbol{\kappa}_{n+1} = \frac{1}{n+1} \begin{bmatrix} n\boldsymbol{\kappa}_n + \vec{b} \\ \kappa_{n+1} \end{bmatrix}, \quad (18)$$

with

$$\kappa_{n+1} = \sum_{i=1}^{n+1} \kappa(\mathbf{x}_{n+1}, \mathbf{x}_i). \quad (19)$$

Computing $\boldsymbol{\kappa}_{n+1}$ as (19) needs to save all the samples $\{\mathbf{x}_i\}_{i=1}^{n+1}$ in memory. For conquering this issue, it can compute as $\boldsymbol{\kappa}_{n+1} = (n+1)\kappa(\mathbf{x}_{n+1}, \mathbf{x}_{n+1})$ by considering an instant estimation. The update of $\boldsymbol{\alpha}_{n+1}$ from $\boldsymbol{\alpha}_n$ is

$$\begin{aligned} \boldsymbol{\alpha}_{n+1} &= \frac{1}{n+1} \begin{bmatrix} n\boldsymbol{\alpha}_n + \mathbf{K}_n^{-1} \mathbf{b} \\ 0 \end{bmatrix} \\ &\quad - \frac{1}{(n+1)(\kappa(\mathbf{x}_{n+1}, \mathbf{x}_{n+1}) - \mathbf{b}^\top \mathbf{K}_n^{-1} \mathbf{b})} \\ &\quad \times \begin{bmatrix} \mathbf{K}_n^{-1} \mathbf{b} \\ 1 \end{bmatrix} (n\mathbf{b}^\top \boldsymbol{\alpha}_n + \mathbf{b}^\top \mathbf{K}_n^{-1} \mathbf{b} - \kappa_{n+1}). \end{aligned} \quad (20)$$

Based on (20), we have an online implementation of the one-class SVM learning phase.

4. Abnormal Events Detection

In an abnormal event detection problem, it is assumed that a set of training frames $\{I_1, \dots, I_n\}$ (the positive class) describing the normal behavior is obtained. The general architectures of abnormal detection are introduced below.

The offline training strategy refers to the case where all the training samples are learnt as one batch, as shown in Figure 3(a). We propose two abnormal detection strategies; the difference between these two strategies is the time when the dictionary is fixed. These two strategies are shown in Figures 3(b) and 3(c). Strategy 1 is shown in Figure 3(b). The training data are learnt one by one. When the training period is finished, the dictionary and the classifier are fixed. Each test datum is classified based on the dictionary. Figure 3(c) illustrates Strategy 2. The training procedure is the same as Strategy 1. But in the testing period, the dictionary is updated if the datum \mathbf{x}_i satisfies the dictionary update condition. The details of these two strategies are explained in the following.

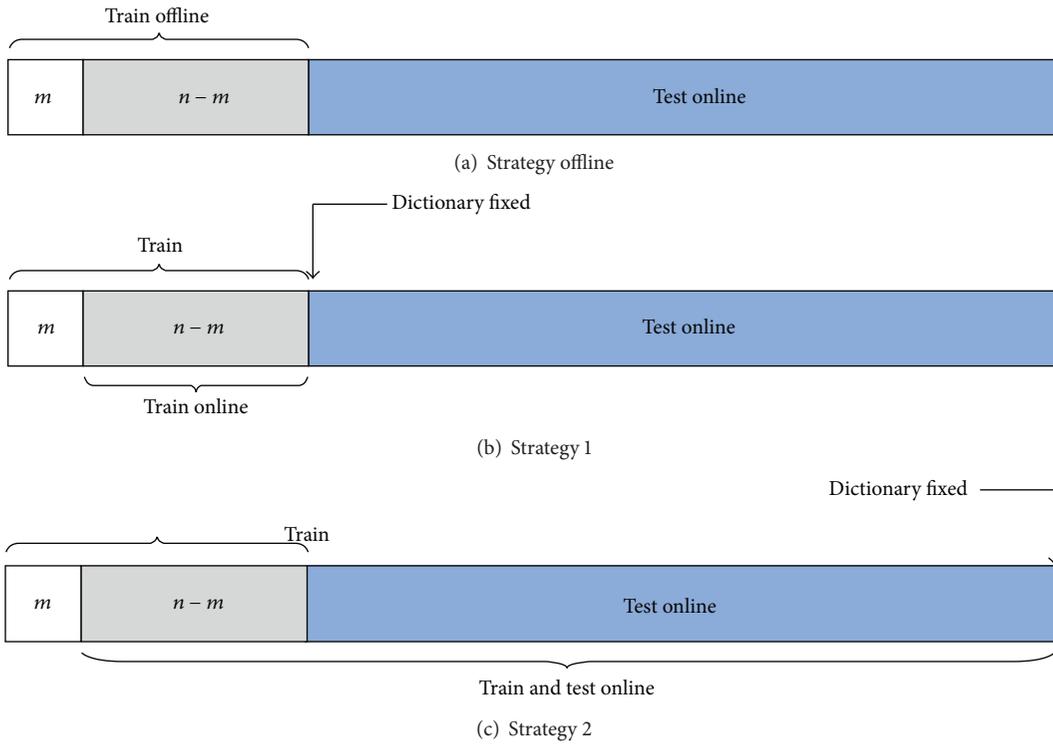


FIGURE 3: Offline and two online abnormal event detection strategies based on one-class SVM. (a) Strategy offline. The training data are learnt as one batch offline. (b) Strategy 1. The dictionary is fixed when all the training data are learnt. (c) Strategy 2. The dictionary continues being updated through the testing period.

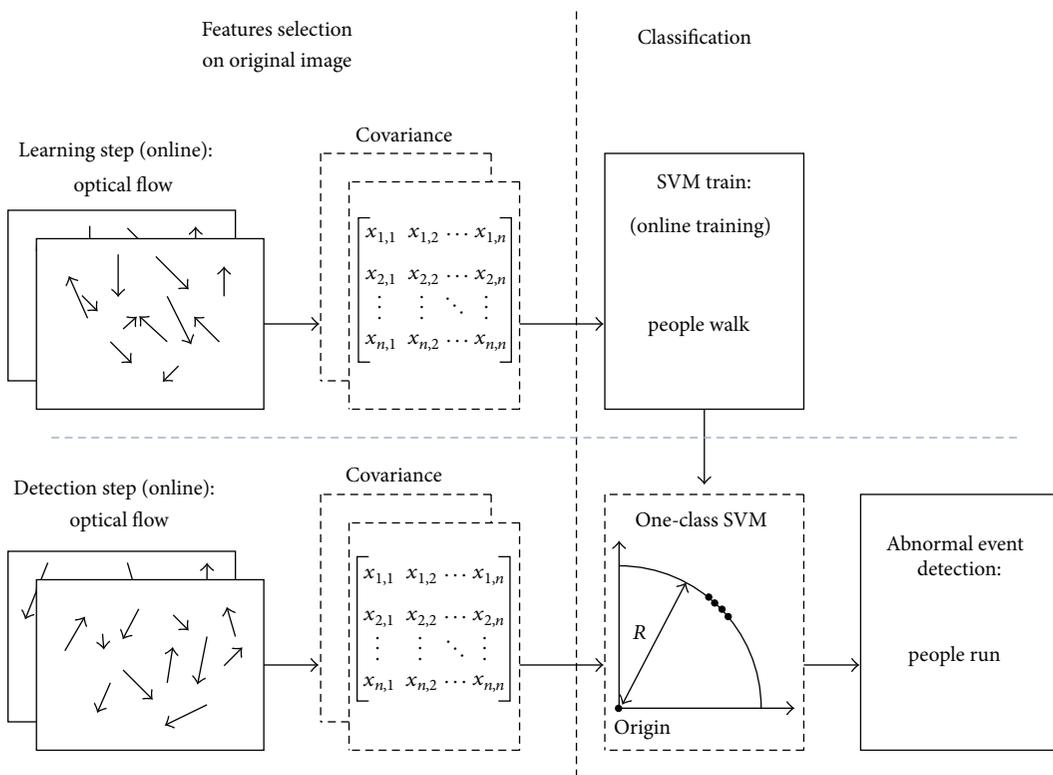


FIGURE 4: Major processing states of the proposed abnormal frame event detection method. The covariance of the frame is computed.

4.1. Strategy 1. In Strategy 1, the dictionary is updated merely through the training period.

Step 1. The first step is calculating the covariance matrix descriptor of training frames based on the image intensity and the optical flow. This step can be generalized as

$$\{(I_1, OP_1), (I_2, OP_2), \dots, (I_n, OP_n)\} \longrightarrow \{C_1, C_2, \dots, C_n\}, \quad (21)$$

where $\{(I_1, OP_1), (I_2, OP_2), \dots, (I_n, OP_n)\}$ are the image intensity and the corresponding optical flow of the 1st to n th frame. $\{C_1, C_2, \dots, C_n\}$ are the covariance matrix descriptors.

Step 2. The second step consists of applying one-class SVM on the small subset of extracted descriptor of the training normal frames to obtain the support vectors. Consider a subset $\{C_i\}_{i=1}^m$, $1 \leq m \ll n$ of data selected from the full training sample set $\{C_i\}_{i=1}^n$, without loss of generality, and assume that the first m examples are chosen. This set of m examples is called dictionary $C_{\mathfrak{D}}$:

$$\{C_1, C_2, \dots, C_m\}, 1 \leq m \ll n \xrightarrow{\text{SVM}} \text{support vector } \{Sp_1, Sp_2, \dots, Sp_o\}, \quad (22)$$

where the set $\{C_1, C_2, \dots, C_m\}$ is the first m covariance matrix descriptors of the training frames; it is the original dictionary $C_{\mathfrak{D}}$. In one-class SVM, the majority of the training samples do not contribute to the definition of the decision function. The entries of a monitory subset of the training samples, $\{Sp_1, Sp_2, \dots, Sp_o\}$, $o \leq m$, are support vectors contributing to the definition of the decision function.

Step 3. After learning the dictionary $C_{\mathfrak{D}}$ which includes the first m , $1 \leq m \ll n$ samples, the training samples $\{C_{m+1}, C_{m+2}, \dots, C_n\}$ are learned online via the technique described in Section 3. This step can be generalized as

$$\{C_{\mathfrak{D}}, C_k\}, m < k \leq n \xrightarrow{\text{SVM}} \text{support vector } \{Sp_1, Sp_2, \dots, Sp_p\}, \quad o \leq p \leq n, \quad (23)$$

$$C_{\mathfrak{D}} := C_{\mathfrak{D}} \cup C_k, \quad \text{if } \epsilon_t \geq \mu_o,$$

where $C_{\mathfrak{D}}$ is the dictionary obtained through Step 2, C_k is a new sample in the remaining training dataset. According to the criterion introduced in Section 3, if the new sample C_k satisfies the dictionary updated condition, it will be included into the dictionary $C_{\mathfrak{D}}$.

Step 4. Based on the dictionary and the classifier obtained from the training frames, the incoming frame sample C_{n+l} is

TABLE 2: AUC of abnormal event detection method of different features F via original SVDD which learns training samples offline, Strategy 1 online one-class SVM, and Strategy 2 online one-class SVM. The biggest value of each method is shown in bold.

Features	Area under ROC		
	Lawn	Indoor	Plaza
Training samples are learned offline			
$F_1 (6 \times 6du)$	0.9426	0.8351	0.9323
$F_2 (6 \times 6dv)$	0.9400	0.8358	0.9321
$F_3 (8 \times 8)$	0.9440	0.8375	0.9359
$F_4 (12 \times 12)$	0.9591	0.8440	0.9580
$F_5 (17 \times 17)$	0.9567	0.8649	0.9649
Strategy 1			
$F_1 (6 \times 6du)$	0.9399	0.8328	0.9343
$F_2 (6 \times 6dv)$	0.9390	0.8355	0.9366
$F_3 (8 \times 8)$	0.9418	0.8377	0.9411
$F_4 (12 \times 12)$	0.9581	0.8457	0.9573
$F_5 (17 \times 17)$	0.9551	0.8628	0.9632
Strategy 2			
$F_1 (6 \times 6du)$	0.9427	0.8237	0.9288
$F_2 (6 \times 6dv)$	0.9370	0.8241	0.9283
$F_3 (8 \times 8)$	0.9430	0.8274	0.9312
$F_4 (12 \times 12)$	0.9605	0.8331	0.9505
$F_5 (17 \times 17)$	0.9601	0.8495	0.9746

classified. The workflow of Strategy 1 is shown in Figure 4 and described by the following equation:

$$f(C_{n+l}) = \text{sgn} \left(R^2 - \sum_{i,j=1}^n \alpha_i \alpha_j \kappa(C_i, C_j) + 2 \sum_i \alpha_i \kappa(C_i, C_{n+l}) - \kappa(C_{n+l}, C_{n+l}) \right) = \begin{cases} 1 & f(C_{n+l}) \geq 0 \\ -1 & f(C_{n+l}) < 0, \end{cases} \quad (24)$$

where C_{n+l} is the covariance matrix descriptor of the $(n+l)$ th frame needed to be classified and C_i and C_j are the samples of the dictionary $C_{\mathfrak{D}}$. "1" corresponds to the normal frame; "-1" corresponds to the abnormal frame.

4.2. Strategy 2. In this strategy, the dictionary is updated through both training and testing periods. The feature extraction step (Step 1) and the online training steps (Steps 2 and 3) are the same as the ones presented in Strategy 1. The testing step is different. The new coming datum which is detected as normal but satisfies dictionary update condition should be included into $C_{\mathfrak{D}}$. The dictionary is needed to be updated through the testing period to include new samples.

Step 4: Strategy 2. If the incoming frame sample C_{n+l} is classified as normal ($f(C_{n+l}) = 1$), the data is checked by the criterion described in Section 3. When the data satisfies

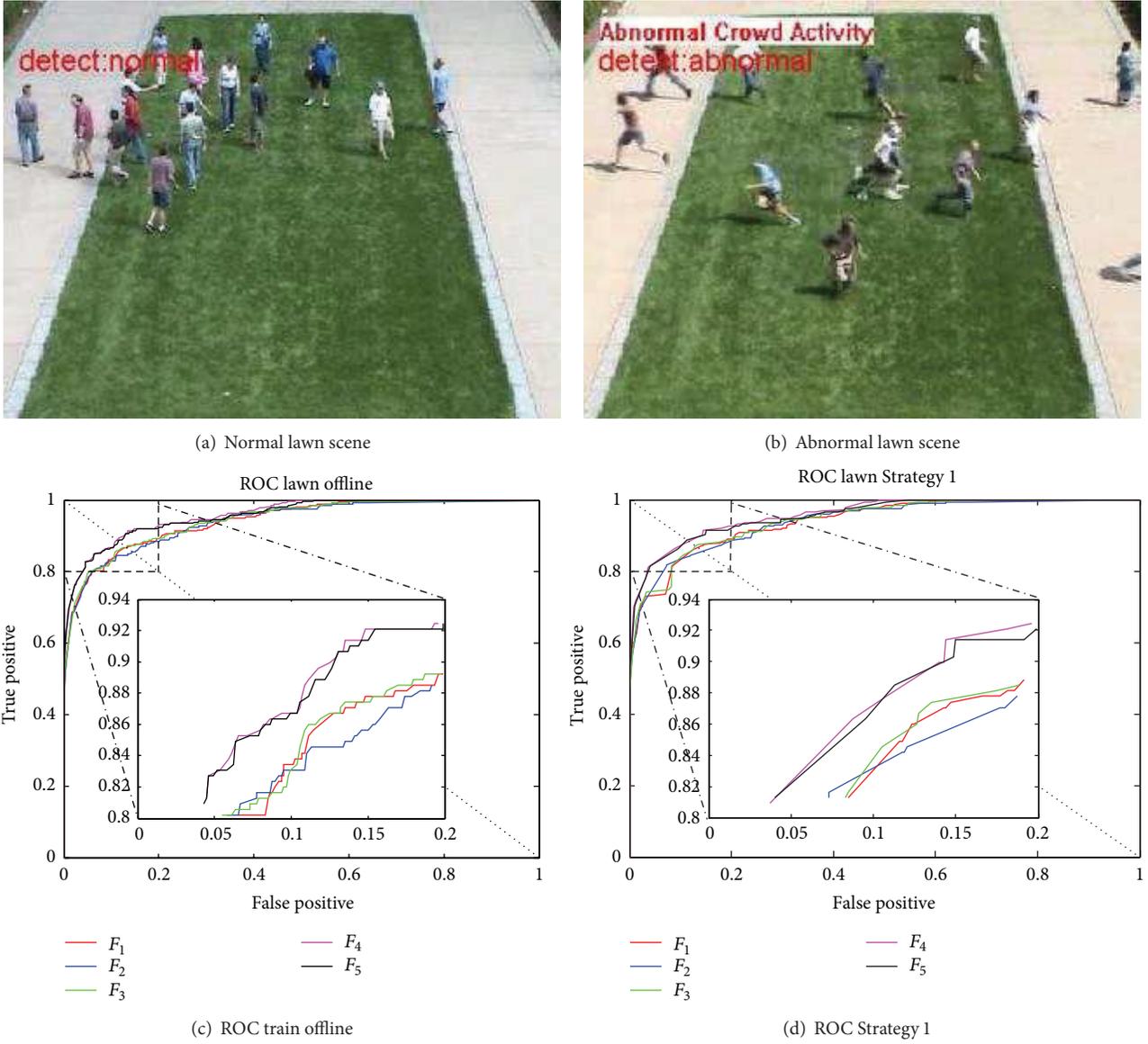


FIGURE 5: Detection results of the lawn scene. (a) The detection result of one normal frame. (b) The detection result of one abnormal panic frame. (c) ROC curve of different features F of the lawn scene results via one-class SVM. All the training samples are learned together offline. The biggest AUC value is 0.9591. (d) ROC curve of different features F results via Strategy 1 online one-class SVM. The biggest AUC value is 0.9581.

the dictionary update criterion, this testing sample will be included into the dictionary. This step can be generalized by the following equation:

$$\begin{aligned}
 f(\mathbf{C}_{n+1}) &= \text{sgn} \left(R^2 - \sum_{i,j=1}^n \alpha_i \alpha_j \kappa(\mathbf{C}_i, \mathbf{C}_j) \right. \\
 &\quad \left. + 2 \sum_i \alpha_i \kappa(\mathbf{C}_i, \mathbf{C}_{n+1}) - \kappa(\mathbf{C}_{n+1}, \mathbf{C}_{n+1}) \right) \\
 &= \begin{cases} 1 & f(\mathbf{C}_{n+1}) \geq 0 \\ -1 & f(\mathbf{C}_{n+1}) < 0. \end{cases} \begin{cases} \epsilon_t \geq \mu_0 \rightarrow \mathbf{C}_{\mathcal{D}} := \mathbf{C}_{\mathcal{D}} \cup \mathbf{C}_{n+1} \\ \epsilon_t < \mu_0 \rightarrow \mathbf{C}_{\mathcal{D}} := \mathbf{C}_{\mathcal{D}} \end{cases}
 \end{aligned} \quad (25)$$

5. Abnormal Detection Result

This section presents the results of experiments conducted to analyze the performance of the proposed method. A competitive performance through both Strategy 1 and Strategy 2 of UMN [31] dataset is presented.

5.1. Abnormal Visual Events Detection: Strategy 1. The results of the proposed abnormal events detection method via Strategy 1 online one-class SVM of UMN [31] dataset are shown below.

The UMN dataset includes eleven video sequences of three different scenes (the lawn, indoor, and plaza) of crowded escape events. The normal samples for training or

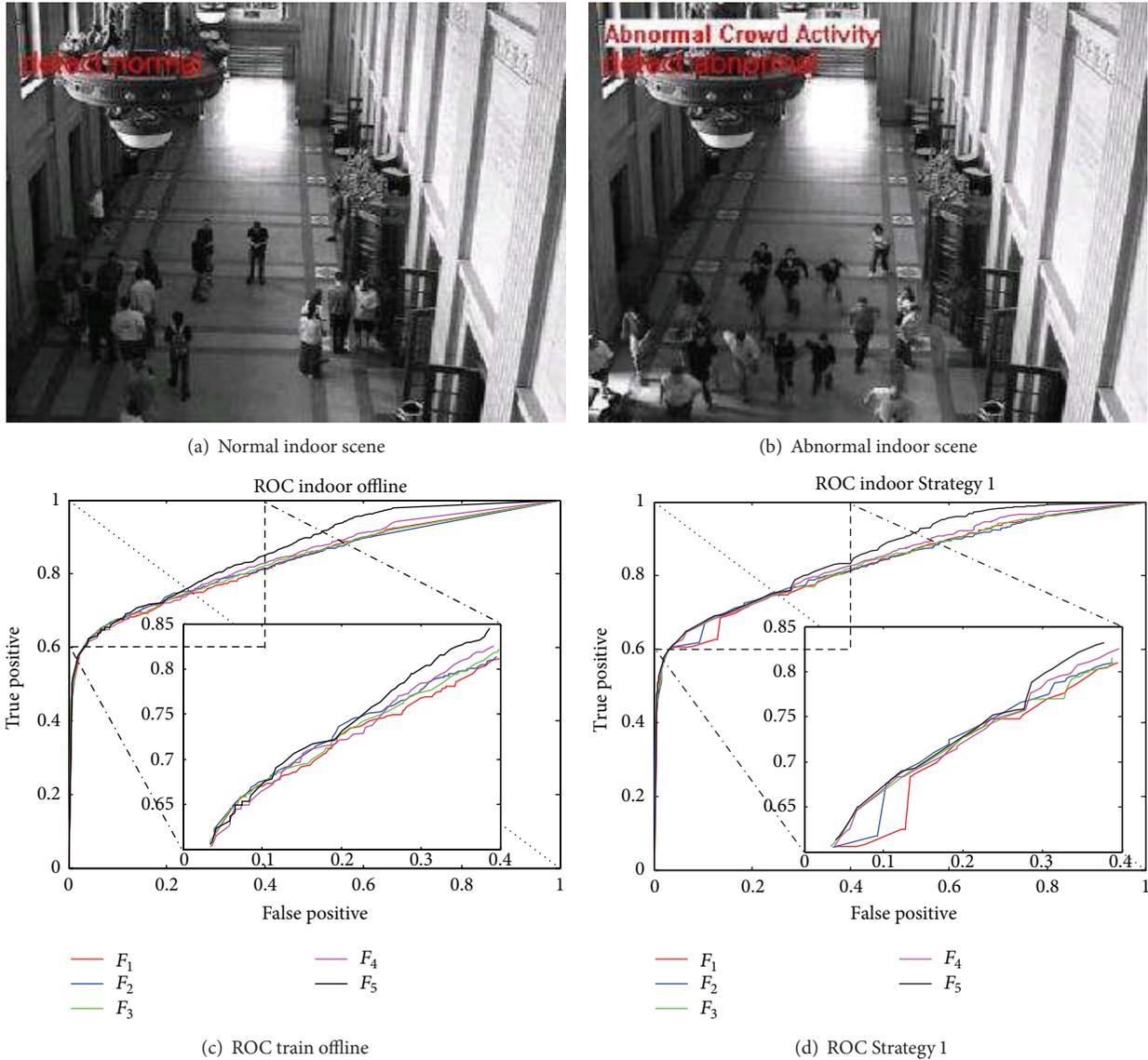


FIGURE 6: Detection results of the indoor scene. (a) The detection result of one normal frame. (b) The detection result of one abnormal panic frame. (c) ROC curve of different features F of the indoor scene results via one-class SVM. All the training samples are learned together offline. The biggest AUC value is 0.8649. (d) ROC curve of different features F results via Strategy 1 online one-class SVM. The biggest AUC value is 0.8628.

TABLE 3: The comparison of our proposed method with the state-of-the-art methods for the abnormal event detection in the UMN dataset.

Method	Area under ROC		
	Lawn	Indoor	Plaza
Social force [16]		0.96	
Optical flow [16]		0.84	
NN [17]		0.93	
SRC [17]	0.995	0.975	0.964
STCOG [18]	0.9362	0.7759	0.9661
COV online (ours)	0.9605	0.8628	0.9746

for normal testing are the frames where the people are walking in different directions. The samples for abnormal testing are the frames where people are running. The detection results of the lawn scene, indoor scene, and plaza scene are shown in Figures 5, 6, and 7, respectively. Gaussian kernel for the Lie Group is used in these three scenes. Different values of σ and penalty factor C are chosen; the area under the ROC curve is shown as a function of these parameters [32]. The results show that taking covariance matrix as descriptor can obtain satisfactory performance for abnormal detection. And also, training the samples online can obtain similarly detection performance as training all the samples offline.

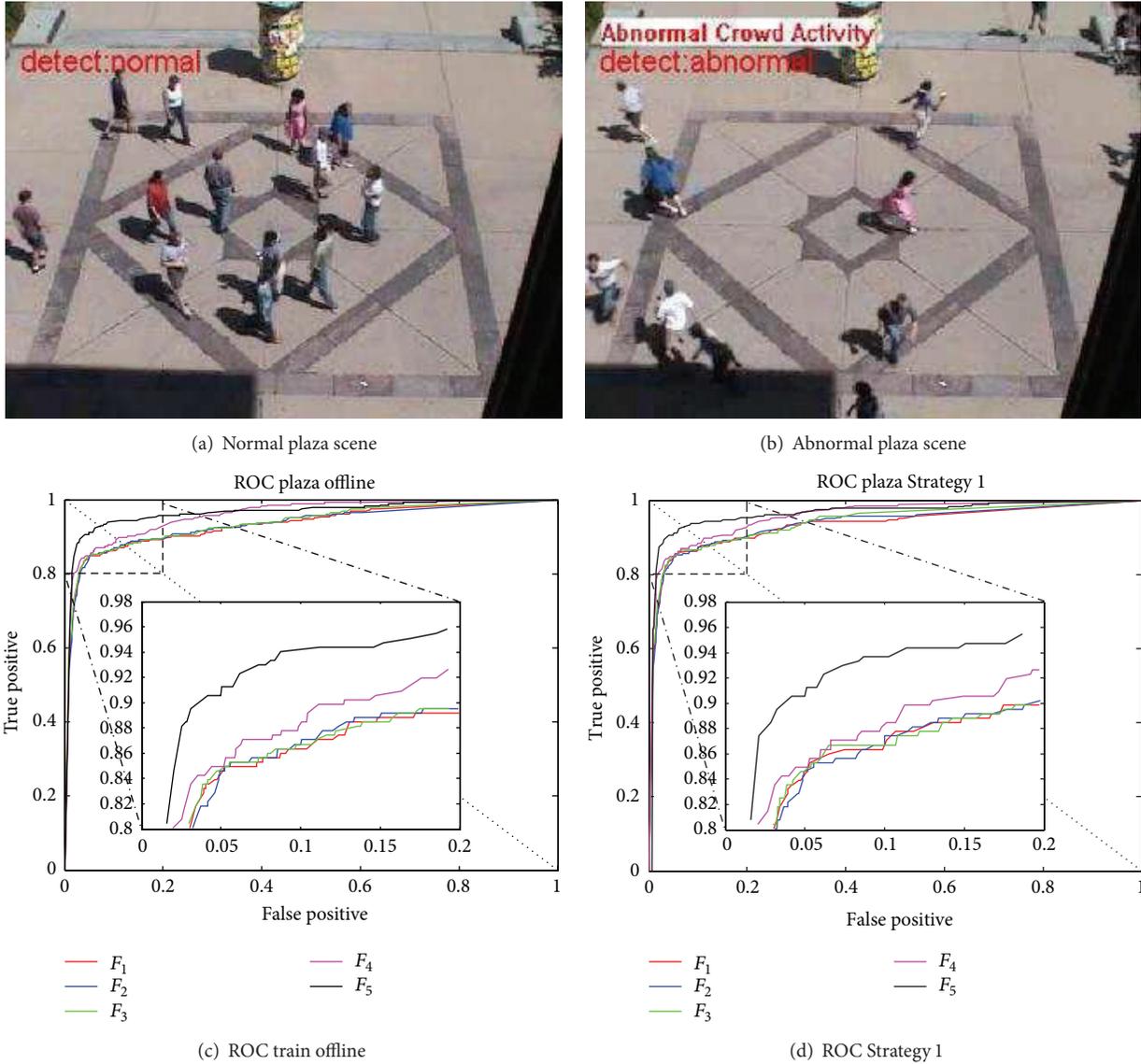


FIGURE 7: Detection results of the plaza scene. (a) The detection result of one normal frame. (b) The detection result of one abnormal panic frame. (c) ROC curve of different features F of the plaza scene results via one-class SVM. All the training samples are learned together offline. The biggest AUC value is 0.9649. (d) ROC curve of different features F results via Strategy 1 online one-class SVM. The biggest AUC value is 0.9632.

Online one-class SVM is appropriate to detect abnormal visual events. There are 1431 frames in the lawn scene, and 480 normal frames are used for training. In the offline strategy, all the 480 frames covariance matrices should be saved in the memory. In Strategy 1, 100 frames covariance matrices are considered as the dictionary firstly. When F_5 -17 \times 17 feature is adopted to construct the covariance descriptor, the variance of Gaussian kernel is $\sigma = 1$, the preset threshold of the criterion is $\mu_0 = 0.5$, the dictionary size increases from 100 to 101, and the maximum accuracy of the detection results is 91.69%. In the indoor scene, there are 2975 normal frames and 1057 abnormal frames. In the plaza scene, there are 1831 normal frames and 286 abnormal frames. The processes of the experiments are similar to the ones of the lawn scene. When feature vector is F_5 -17 \times 17, $\sigma = 1$, $\mu_0 = 0.5$, the dictionary

size of these two scenes remain 100. The online strategy keeps the memory size almost unchanged when the size of training dataset increases.

5.2. Abnormal Visual Events Detection: Strategy 2. The results of the abnormal event detection method via Strategy 2 of UMN dataset are shown as follows. In the experiment process of the lawn scene, 100 normal samples from the training samples are learnt firstly, and then the other 380 training data are learnt online one by one. After these two training steps, we can obtain the basic dictionary from the training samples and also the classifier. In the following testing step, the dictionary is updated if the sample satisfies the dictionary update criterion. When a new sample is coming, it is firstly detected by the previous classifier. If it is classified as anomaly,

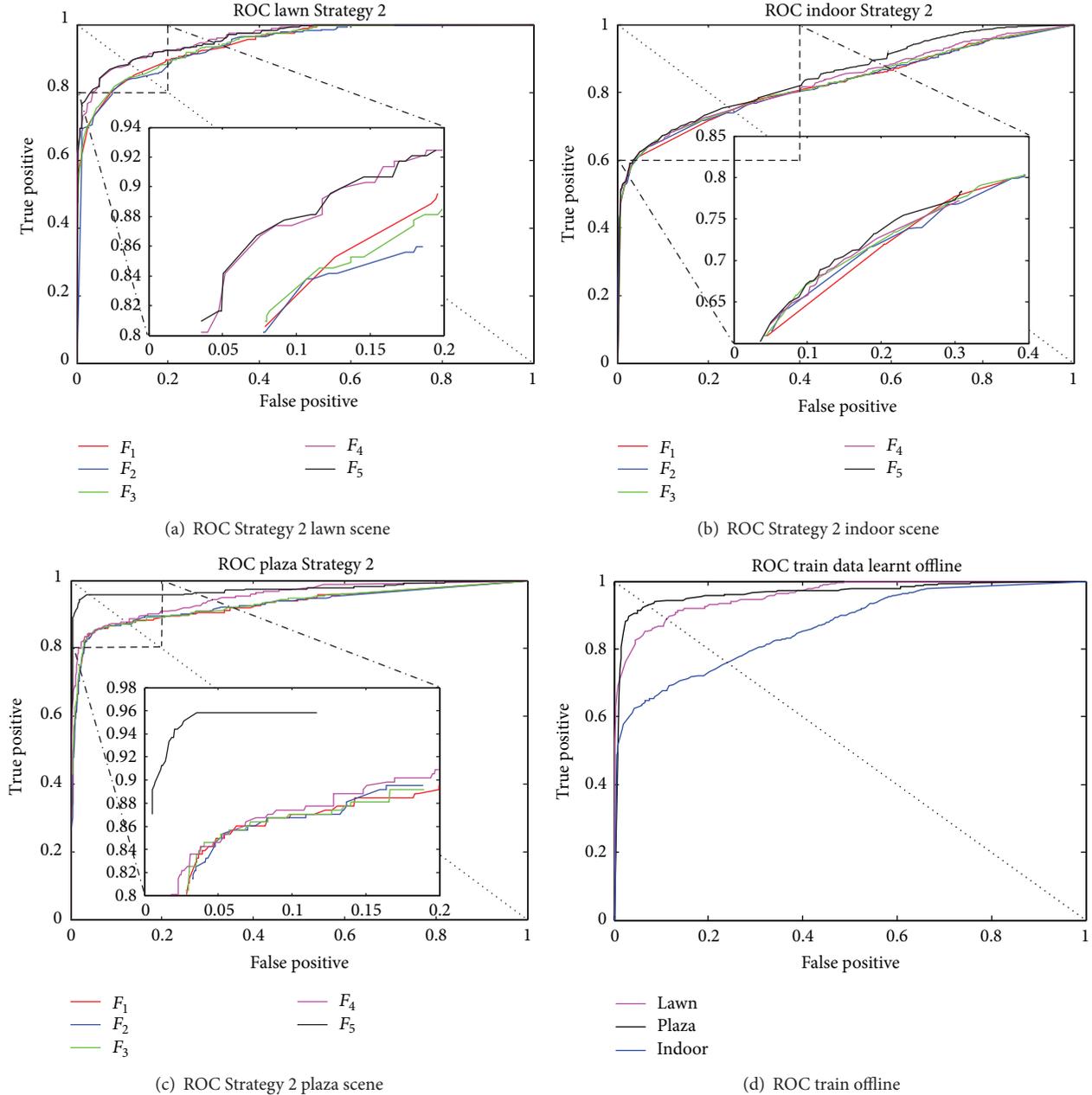


FIGURE 8: ROC curve of UMN dataset. (a) ROC curve of different features F results via Strategy 2 of the lawn scene. The biggest AUC value is 0.9605. (b) Strategy 2 results of the indoor scene. The biggest AUC value is 0.8495. (c) Strategy 2 results of the plaza scene. The biggest AUC value is 0.9746. (d) The ROC curve of best performance of the lawn, indoor, and plaza scene when the training samples are learnt offline. The biggest AUC values of the lawn, indoor, and plaza are 0.9591, 0.8649 and 0.9649.

the dictionary and the classifier are not changed. Otherwise, if the sample is classified as a normal one, the sparse criterion introduced in Section 3 is used to check the correlation between the earlier dictionary and this new datum. It will be included into the dictionary when it satisfied the update condition. The dictionary will be updated through the whole testing period. The other two scenes, the indoor and plaza scene, are handled by the same methods. When F_5 - 17×17 feature is adopted, the variance of the Gaussian kernel is $\sigma = 1$, and the preset threshold of the criterion is $\mu_0 = 0.5$, and the dictionary size of the lawn, indoor, and plaza scene

is increased from 100 to 106, 102, and 102, respectively. The ROC curve of detection results of these three scenes is shown in Figures 8(a), 8(b), and 8(c). Besides the merit of saving memory of Strategy 1, Strategy 2 also has the advantage of adaptation to the long duration sequence.

The results performances of offline strategy, Strategy 1, and Strategy 2 are shown in Table 2. The performances of these two strategies results are similar to that of the results when all training samples are learnt together. When F_4 (12×12) or F_5 (17×17) are chosen as the features to form covariance matrix descriptor, the results have the best performance.

These two features are more abundant to include movement and intensity information.

The result performances of the covariance matrix descriptor based online one-class SVM method and the state-of-the-art methods are shown in Table 3. The covariance matrix based online abnormal frame detection method obtains competitive performance. In general, our method is better than others except sparse reconstruction cost (SRC) [17] in lawn scene and indoor scene. In that paper, multiscale HOF is taken as a feature, and a testing sample is classified by its sparse reconstructor cost, through a weighted linear reconstruction of the overcomplete normal basis set. But computation of the HOF might takes more time than calculating covariance. By adopting the integral image [15], the covariance matrix descriptor of the subimage can be computed conveniently. So the covariance descriptor can be appropriately used to analyze the partial movement.

6. Conclusions

A method for the abnormal event detection of the frame is proposed. The method consists of covariance matrix descriptor encoding the movement features, and the online nonlinear one-class SVM classification method. We have developed two nonlinear one-class SVM based abnormal event detection techniques that update the normal models of the surveillance video data in an online framework. The proposed algorithm has been tested on a video dataset yielding successful results to detect abnormal events.

Acknowledgments

This work is partially supported by China Scholarship Council of Chinese Government, SURECAP CPER Project funded by Région Champagne-Ardenne, and CAPSEC CRCA FEDER platform.

References

- [1] F. Jiang, J. Yuan, S. A. Tsafaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Computer Vision and Image Understanding*, vol. 115, no. 3, pp. 323–333, 2011.
- [2] C. Piciarelli, C. Micheloni, and G. L. Foresti, "Trajectory-based anomalous event detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1544–1554, 2008.
- [3] C. Piciarelli and G. L. Foresti, "On-line trajectory clustering for anomalous events detection," *Pattern Recognition Letters*, vol. 27, no. 15, pp. 1835–1842, 2006.
- [4] T. Xiang and S. Gong, "Incremental and adaptive abnormal behaviour detection," *Computer Vision and Image Understanding*, vol. 111, no. 1, pp. 59–73, 2008.
- [5] S. Gong and T. Xiang, "Recognition of group activities using dynamic probabilistic networks," in *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV '03)*, pp. 742–749, October 2003.
- [6] J. Kim and K. Grauman, "Observe locally, infer globally: a space-time MRF for detecting abnormal activities with incremental updates," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 2921–2928, Miami, Fla, USA, June 2009.
- [7] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, 2008.
- [8] O. Boiman and M. Irani, "Detecting irregularities in images and in video," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 17–31, 2007.
- [9] X. Zhu, Z. Liu, and J. Zhang, "Human activity clustering for online anomaly detection," *Journal of Computers*, vol. 6, no. 6, pp. 1071–1079, 2011.
- [10] X. Zhu and Z. Liu, "Human behavior clustering for anomaly detection," *Frontiers of Computer Science in China*, vol. 5, no. 3, pp. 279–289, 2011.
- [11] Y. Benezeth, P. M. Jodoin, and V. Saligrama, "Abnormality detection using low-level co-occurring events," *Pattern Recognition Letters*, vol. 32, no. 3, pp. 423–431, 2011.
- [12] Y. Benezeth, P. M. Jodoin, V. Saligrama, and C. Rosenberger, "Abnormal events detection based on spatio-temporal co-occurrences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 2458–2465, Miami, Fla, USA, June 2009.
- [13] A. Mittal, A. Monnet, and N. Paragios, "Scene modeling and change detection in dynamic scenes: a subspace approach," *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 63–79, 2009.
- [14] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.
- [15] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: a fast descriptor for detection and classification," in *Computer Vision—ECCV 2006*, vol. 3952 of *Lecture Notes in Computer Science*, pp. 589–600, Springer, New York, NY, USA, 2006.
- [16] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 935–942, Miami, Fla, USA, June 2009.
- [17] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 3449–3456, Providence, RI, USA, June 2011.
- [18] Y. Shi, Y. Gao, and R. Wang, "Real-time abnormal event detection in complicated scenes," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 3653–3656, Istanbul, Turkey, August 2010.
- [19] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, MIT Press, Boston, Mass, USA, 2002.
- [20] B. Hall, *Lie Groups, Lie Algebras, And Representations: An Elementary Introduction*, vol. 222, Springer, New York, NY, USA, 2003.
- [21] C. Gao, F. Li, and C. Shen, "Research on lie group kernel learning algorithm," *Journal of Frontiers of Computer Science and Technology*, vol. 6, pp. 1026–1038, 2012.
- [22] V. N. Vapnik and A. Lerner, "Pattern recognition using generalized portrait method," *Automation and Remote Control*, vol. 24, pp. 774–780, 1963.
- [23] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 2000.
- [24] D. Tax, *One-class classification [Ph.D. thesis]*, Delft University of Technology, 2001.

- [25] D. M. Tax and R. P. Duin, "Data domain description using support vectors," in *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN '99)*, vol. 99, pp. 251–256, 1999.
- [26] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [27] Z. Noumir, P. Honeine, and C. Richard, "Online one-class machines based on the coherence criterion," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO '12)*, pp. 664–668, 2012.
- [28] Z. Noumir, P. Honeine, and C. Richard, "One-class machines based on the coherence criterion," in *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP '12)*, pp. 600–603, 2012.
- [29] P. Honeine, "Online kernel principal component analysis: a reducedorder model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1814–1826, 2012.
- [30] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, 2009.
- [31] UMN, "Unusual crowd activity dataset of university of Minnesota," 2006, http://mha.cs.umn.edu/proj_events.shtml.
- [32] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

