2008

Abnormal behavior detection using a multi-modal stochastic learning approach

Philippa M. Lyons-Wall
University of Wollongong, philippa@uow.edu.au

Abdesselam Bouzerdoum
University of Wollongong, bouzer@uow.edu.au

Son Lam Phung
University of Wollongong, phung@uow.edu.au

Azeddine Beghdadi
Institut Galilée, beghdadi@galilee.univ-paris13.fr

Abnormal behavior detection using a multi-modal stochastic learning approach

Abstract

This paper presents a new approach to trajectory-based abnormal behavior detection (ABD). While existing techniques include position in the feature vector, we propose to estimate the probability distribution locally at each position, hence reducing the dimensionality of the feature vector. Local information derived from accumulated knowledge for a particular position is integrated in the distribution enabling context-based decision for ABD. A stochastic competitive learning algorithm is employed to estimate the local distributions of the feature vector and the location of the distribution modes. The proposed algorithm is tested on the detection of driving under the influence of alcohol. The performance of the new algorithm is evaluated on synthetic data. First the local stochastic learning algorithm is compared to its global variant. Then it is compared to the Kohonen self organizing feature maps. In both cases, the proposed algorithm achieves higher detection rates (at the same false alarm rate) with fewer clusters.

Disciplines

Physical Sciences and Mathematics

Publication Details

P. L. M. Bouttefroy, A. Bouzerdoum, S. Lam. Phung & A. Beghdadi, "Abnormal behavior detection using a multi-modal stochastic learning approach," in International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 2008, pp. 121-126.

Abnormal Behavior Detection using a Multi-Modal Stochastic Learning Approach

P. L. M. Bouttefroy, A. Bouzerdoum, S. L. Phung
School of Electrical, Computer and Telecommunications Engineering
University of Wollongong, Wollongong, NSW, Australia
{philippe, a.bouzerdoum, phung}@uow.edu.au

A. Beghdadi
L2TI, Institut Galilée
Université Paris 13, Villetaneuse, France
beghdadi@galilee.univ-paris13.fr

***Abstract*—This paper presents a new approach to trajectory-based Abnormal Behavior Detection (ABD). While existing techniques include position in the feature vector, we propose to estimate the probability distribution locally at each position, hence reducing the dimensionality of the feature vector. Local information derived from accumulated knowledge for a particular position is integrated in the distribution enabling context-based decision for ABD. A stochastic competitive learning algorithm is employed to estimate the local distributions of the feature vector and the location of the distribution modes. The proposed algorithm is tested on the detection of driving under the influence of alcohol. The performance of the new algorithm is evaluated on synthetic data. First the local stochastic learning algorithm is compared to its global variant. Then it is compared to the Kohonen self organizing feature maps. In both cases, the proposed algorithm achieves higher detection rates (at the same false alarm rate) with fewer clusters.**

I. Introduction

The last decade witnessed a rise in the development of abnormal behavior detection (ABD) systems due to the increasing demand in visual surveillance and security. The development of ABD systems has been possible due to the recent ability of computers to process information from videos in real-time. Even though ABD remains an open problem, the inherent outcomes are already tremendous. Robust ABD will enable the automatic detection of abnormal events and the instantaneous notification of the relevant authority. It would result in a dramatic reduction in the need for human effort to analyze video-surveillance sequences, for instance. However, robust ABD still encounters a number of challenging problems because it is a high-level process requiring the integration of various techniques such as feature selection, trajectory modeling, dimensionality reduction and density estimation. Most of these fields are still active areas of research; global optimal solutions are yet to be found for many of the aforementioned problems.

Abnormal behavior detection can be classified in two different categories. The first one is concerned with the characterization of the body posture. The second category is based upon the analysis of the object elementary features such as position, color, displacement, object size, and so on. The algorithm proposed here falls in the latter category. Abnormal behavior detection through feature-based analysis aims at estimating the probability distribution of the feature vector representing the object. In a *global* approach, the probability distribution of the feature vector is estimated over the entire feature space. The feature-based analysis traditionally detects abnormal behavior based on the distribution of the feature vector after a learning phase. Different ABD algorithms differ in the way they learn the probability distribution. Hidden Markov models are typically used to track and detect abnormal behavior [2], [6], [12]. Neural networks have also been used for ABD with some success [3]. In particular, self organizing maps (SOMs) have attracted attention due to their ability to preserve the topology of the input data [5], [13]. Johnson and Hogg have also proposed to model the probability distribution of a feature vector composed of the position and the vector flow of the object using a competitive neural network [8] and a global Gaussian mixture model [9].

Unlike the aforementioned approaches, the algorithm developed here proposes to further hone the characteristics of the feature vector and, in particular, to differentiate between *local* and *behavioral* features. Local features provide information that is characteristic of the object in a local environment. For instance, the position of the object is a local feature. With a global model, the estimation of the local feature distribution does not provide optimal results because the information depends on the local context. Nevertheless, local features are of crucial importance to explain the local behavior of an object. Therefore, they are used to index the distribution. A map of local probability distributions for the behavioral features is thus generated. The map is accessed through the local features (as indexes). The benefits are twofold: first, the dimensionality of the feature vector is reduced; second, the error in the distribution estimate is reduced because the local features are not estimated by the global model.

The paper is organized as follows. Section II introduces the position-based modeling of the feature vector. Section III describes the integration of the neighborhood information in the model. Section IV presents the stochastic update of the local distribution. Finally, Section V presents the experimental setting and results before concluding in Section VI.

II. Mixture Model of the Feature Vector Density

Trajectory-based ABD aims to estimate the distribution of the feature vector $\mathbf{X}$ over time. Markov chains provide a convenient framework to estimate such a distribution when knowledge about the state for the previous time step is known.

The problem, modeled in mathematical terms, is equivalent to determining the conditional probability $P(\mathbf{X}_t|\mathbf{X}_{t-1})$ where $\mathbf{X}_t$ and $\mathbf{X}_{t-1}$ are the feature vectors at times t and $t-1$, respectively [10], [7]. The contribution of this paper is to estimate the feature vector probability distribution characterizing an object at time t given its feature vector at time $t-1$ for a particular local feature. To this end, we split the feature vector $\mathbf{X}$ into two components: a local component $\mathbf{S}$ representing the coordinates and a behavioral component φ representing the behavior of the object. Thus, the feature vector can be rewritten as $\mathbf{X} = \{\mathbf{S}, \varphi\}$.

A. Dimensionality Reduction via Spatial Feature Indexation

Let us introduce the general framework of abnormal behavior detection before discussing the motivations for dimensionality reduction. We suppose that the problem of ABD can be described as a Markov chain; one can recursively determine the distribution of the feature vector with conditional probabilities. It can thus be assumed that the distribution of the feature vector at time t is described by the probability density $p(\mathbf{X}_t|\mathbf{X}_{t-1})$. The spatial marginal distribution $p(\mathbf{S}|\mathbf{X}_{t-1})$ and the behavioral marginal distribution $p(\varphi|\mathbf{X}_{t-1})$ are defined as:

$$p(\mathbf{S}|\mathbf{X}_{t-1}) = \int_{D_\varphi} p(\mathbf{X}_t|\mathbf{X}_{t-1})\, d\varphi\,; \quad \text{and} \tag{1}$$

$$p(\varphi|\mathbf{X}_{t-1}) = \int_{D_\mathbf{S}} p(\mathbf{X}_t|\mathbf{X}_{t-1})\, d\mathbf{S}\,. \tag{2}$$

where $D_\mathbf{S}$ and D_φ are the respective definition domains of the components $\mathbf{S}$ and φ. The marginal distributions provide a representation of the spatial and behavioral component spans over their respective subspaces, namely $D_\mathbf{S}$ and D_φ.

The motivation underlying the distinction between behavioral features and spatial features is the sparsity of the spatial component of the dataset. Indeed, the amount of data available is not homogeneously distributed over the feature space. Typically, traffic sequences present high density of objects in certain areas of the scene while others are deserted. Consequently, the spatial marginal distribution $p(\mathbf{S}|\mathbf{X}_{t-1})$ is difficult to approximate and the error in the estimation is large. In contrast, the behavioral marginal distribution $p(\varphi|\mathbf{X}_{t-1})$ is usually dense and can be estimated accurately. Figure 1 shows an example of such marginal distributions. The set of object position is sparse and clearly follows specific patterns (called routes) while its behavior, represented by the vector flow, is dense. As a result, the spatial component of the feature vector is accountable for most of the estimation error on the distribution. Nevertheless, the position of the object plays an important role in the analysis of the behavior because it provides information on the local environment and, thus, has to be considered for abnormal behavior detection; the behavioral distribution is not sufficient to discriminate abnormal behavior. Therefore, we propose to reduce the estimation of the feature vector distribution $p(\mathbf{X}_t|\mathbf{X}_{t-1})$ to the behavioral component $p(\varphi|\mathbf{X}_{t-1})$ and use the spatial component $\mathbf{S}$ to index the local behavioral distributions. This results in a local characterization of the behavior. The advantage is twofold. First, the dimensionality of the feature space is reduced, avoiding the so-called curse of dimensionality. Second, the behavior is locally analyzed, enabling abnormal behavior detection based on contextual information.

B. Local Probability Density Estimation

Because the probability density estimate $p(\mathbf{X}_t|\mathbf{X}_{t-1})$ is derived from local features, the coordinates of the object are not included in the feature vector $\mathbf{X}$ but are rather used to index the probability density. Let us denote by $\mathbf{S}$ the coordinate component and φ the behavioral component (e.g. size, color, shape, etc.) of the feature vector $\mathbf{X}$. The state of the object at time t can be expressed as follows:

$$\mathbf{X}_t = \mathbf{X}_{t-1} + f_\mathbf{S}(\varphi)\,, \tag{3}$$

where $f_\mathbf{S}(\varphi)$ is a local function of the behavioral component. Such a function is difficult to estimate directly because there is no knowledge of behaviors in the scene. Instead, we propose to approximate $f_\mathbf{S}(\varphi)$ by sampling from a local conditional *probability density function* (pdf):

$$p(\mathbf{X}_t|\mathbf{X}_{t-1}) \propto p_\mathbf{S}(\varphi|\mathbf{X}_{t-1})\,. \tag{4}$$

The probability density function $p_\mathbf{S}(\varphi|\mathbf{X}_{t-1})$ is recursively derived from the knowledge accumulated over time. We propose to model the probability density of the feature vector with a fixed number of clusters K representing the modes of the density. Thus, the pdf of the feature vector can be expressed as

$$p_\mathbf{S}(\varphi|X_{t-1}) = \gamma \sum_{k=1}^{K} f_k(\varphi; \mu_k)\,. \tag{5}$$

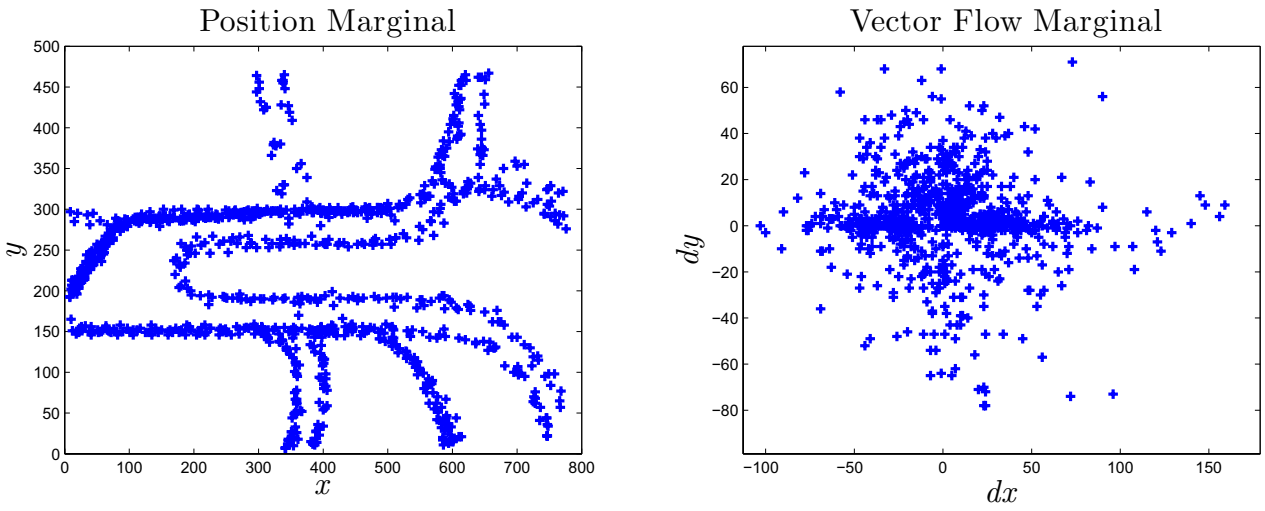
where γ is a normalizing constant. Here, the function $f_k(\varphi; \mu_\mathbf{k})$ is defined as follows:

$$f_k(\varphi; \mu_\mathbf{k}) = \frac{w_k}{1 + \exp(||\varphi - \mu_\mathbf{k}||)}\,, \tag{6}$$

where $||\varphi - \mu_\mathbf{k}||$ denotes the Euclidian distance between the vectors φ and $\mu_\mathbf{k}$, and w_k are the weights. The parameters $\mu_\mathbf{k}$ approximate the position of the modes in the density.

III. Integration of Local Information

The model defined in (5) is a temporal estimate of the probability density function. It represents the probability distribution of the behavioral component for a particular local component. Because objects in the same neighborhood tend to have the same behavior, the feature vectors at neighboring locations are highly correlated. Based on this hypothesis, we propose an algorithm that integrates information from a local neighborhood in order to estimate the modes of the distribution. The model and more specifically the modes μ_k of the local density are thus temporally and spatially updated as new information arrives. Let us assume that the probability



(a) (b)

Fig. 1. Example of marginal distributions for a feature vector of dimension 4. (a) Marginal distribution of the spatial component $\mathbf{S}$, i.e. position of objects in the scene; (b) Marginal distribution of the behavioral component φ, representing the distribution of the displacement of objects in the scene (vector flow).

density $p_{\mathbf{S}}(\varphi|\mathbf{X}_{t-1})$ is conditioned on two, and only two, independent events at time $t-1$, e.g. two objects in motion. The first event is represented by the feature vector $\mathbf{X}^1_{t-1}$ with local feature $\mathbf{S}$, and the second event by $\mathbf{X}^2_{t-1}$ with local feature $\mathbf{S}+\delta_{\mathbf{S}}$, in the neighborhood of $\mathbf{X}^1_{t-1}$. The probability density $p_{\mathbf{S}}(\varphi|\mathbf{X}_{t-1})$ is then given by Eq. (7).

$$\begin{aligned} p_{\mathbf{S}}(\varphi|X_{t-1}) &= p_{\mathbf{S}}(\varphi|X^1_{t-1} \cup X^2_{t-1}) \\ &= \frac{p_{\mathbf{S}}(\varphi)}{p_{\mathbf{S}}(X_{t-1})}\left[p_{\mathbf{S}}(X^1_{t-1}|\varphi) + p_{\mathbf{S}}(X^2_{t-1}|\varphi)\right] \qquad (7) \\ &= \frac{p_{\mathbf{S}}(X^1_{t-1})p_{\mathbf{S}}(\varphi|X^1_{t-1}) + p_{\mathbf{S}}(X^2_{t-1})p_{\mathbf{S}}(\varphi|X^2_{t-1})}{p_{\mathbf{S}}(X_{t-1})} \end{aligned}$$

The pdf $p_{\mathbf{S}}(\varphi|\mathbf{X}^2_{t-1})$ is not known for the local feature $\mathbf{S}$ as the event $\mathbf{X}^2_{t-1}$ occurs at location $\mathbf{S}+\delta_{\mathbf{S}}$, only the probability density $p_{\mathbf{S}+\delta_{\mathbf{S}}}(\varphi|\mathbf{X}^2_{t-1})$ is. Nevertheless, the two probabilities are related by

$$p_{\mathbf{S}}(\varphi|\mathbf{X}^2_{t-1}) = \xi(\delta_{\mathbf{S}})p_{(\mathbf{S}+\delta_{\mathbf{S}})}(\varphi|\mathbf{X}^2_{t-1})\,, \qquad (8)$$

where $\xi(\delta_{\mathbf{S}})$ is an unknown function. We assume ξ follows the definition of a kernel, as stated in [4], and, in particular, its value decreases with the distance from the origin. We choose the isotropic Gaussian to model the function $\xi(\delta_{\mathbf{S}})$:

$$\xi(\delta_{\mathbf{S}}) = \frac{1}{\sqrt{2\pi}\sigma}e^{-||\delta_{\mathbf{S}}||^2/2\sigma^2}\,, \qquad (9)$$

where $||.||$ denotes the Euclidean norm. Because the weights w_k are normalized, the probability mass function $p_{\mathbf{S}}(\mathbf{X}_{t-1})$ is equal to 1. It follows from Eqs. (5), (7) and (8) that

$$p_{\mathbf{S}}(\varphi|X_{t-1}) = \gamma\sum_{k=1}^{K}\left[\xi(0)f_k(\varphi;\mu^1_k)+\xi(\delta_{\mathbf{S}})f_k(\varphi;\mu^2_k)\right]. \qquad (10)$$

Equation (10) can be generalized to N ($N \in \mathbb{N}$) independent events. Let us denote $l(N)$ the set of local features $\{\mathbf{S}_0, ..., \mathbf{S}_N\}$ for the N events. The pdf $p_{\mathbf{S}}(\varphi|\mathbf{X}_{t-1})$ is then given by

$$p_{\mathbf{S}}(\varphi|X_{t-1}) = \gamma\sum_{i=1}^{N}\sum_{k=1}^{K}\xi(\delta_{\mathbf{S}_i})f_k(\varphi;\mu^i_k)\,, \qquad (11)$$

where the variable μ^i_k denotes the k^{th} cluster center for the i^{th} local feature $\mathbf{S}_i$.

IV. Updating the Clusters with Stochastic Competitive Learning

The probability distribution of an event is estimated through the spatio-temporal mixture as defined in Eq. (11), making it dependent on its neighborhood. In turn, the event is also used to update the parameters of all the distributions in a small neighborhood. In particular, the density $p_{\mathbf{S}}(\varphi|X_{t-1})$ is temporally adjusted via the update of the modes $\mu_{\mathbf{k}}$. The clusters for each spatial component $\mathbf{S}_i$ are updated following the algorithm developed by Bouzerdoum [1]. A stochastic procedure is adopted to allocate the clusters in the feature space. The competitive learning process is thus defined as follows. Consider the cluster center affinity y^i_k (for $k = 1..K$) to the input feature vector:

$$y^i_k = f_k(\varphi;\mu^i_k) + r^i_k, \qquad (12)$$

where r_k^i is a centered normal random variable with standard deviation σ_k^i. The winning cluster l^i is determined by competitive learning as:

$$l^i = \arg\max_k (y_k^i)\,. \tag{13}$$

The parameter μ_l^i of the winning cluster for each spatial component $\mathbf{S}_i$ is updated by a low-pass first-order filter with learning rate α where $0 < \alpha < 1$:

$$\mu_l^i \leftarrow \left[1 - \alpha\,\xi(\delta_{\mathbf{S}_i})\right]\mu_l^i + \alpha\,\xi(\delta_{\mathbf{S}_i})\,\varphi\,, \tag{14}$$

As the cluster learns, the standard deviation of the random variable r_k^i in (12) is reduced to allow convergence. A counter n_l^i is incremented with the magnitude of the kernel for the spatial component $\mathbf{S}_i$ by

$$n_l^i \leftarrow n_l^i + \xi(\delta_{\mathbf{S}_i})\,, \tag{15}$$

and the standard deviation is updated as follows:

$$\sigma_l^i = \sigma_0 / n_l^i\,. \tag{16}$$

The weight w_l of the winning cluster is incremented by $\xi(\delta_{\mathbf{S}_i})$ and all weights are normalized. The "shaking" process introduced in the clustering algorithm improves the convergence of the cluster center to the mode of the distribution. Indeed, because of the on-line nature of the learning algorithm, the initialization of the center value is critical. For example, if the center is initialized on an outlier, a standard learning algorithm may not converge to a relevant mode of the distribution. It should also be noted that if the number of outliers increases, the stochastic algorithm performs better in its ability to find the cluster center [1]. In such a case, the variance (σ_l^i) rate of decrease can be lowered to allow more shaking in the learning phase. This provides a better convergence of the parameters μ_l^i to the modes; on the other hand, the convergence will be slower. With the randomness introduced by the stochastic clustering algorithm, the cluster center is shifted around until the number of samples is large enough for an accurate estimation of the mode position.

V. Abnormal Behavior Detection

The estimation of the probability density $p_{\mathbf{S}}(\varphi|X_{t-1})$ is used to determine whether or not a behavior represented by the feature vector φ is abnormal. The higher the probability is, the higher are the chances for an object to have a normal behavior. Indeed, the probability is high when the feature vector fits well the model, i.e. when the object behavior is predictable. On the contrary, low probability means the behavior is unpredictable, hence considered as abnormal. In the proposed model, this is equivalent to measuring the Euclidian distance between the winning cluster as defined in (13) and the feature vector. The behavior is considered as abnormal if the distance is greater than a threshold T and normal otherwise:

$$\begin{cases} ||\varphi - \mu_l|| \leq T & \rightarrow \text{“}normal\text{”}\,; \\ ||\varphi - \mu_l|| > T & \rightarrow \text{“}abnormal\text{”}\,. \end{cases} \tag{17}$$

Each feature vector is thus classified as representing a normal or an abnormal behavior.

A. Experimental Setup

The system is tested on synthetic data modeling the behavior of drivers under the influence of alcohol as abnormal behavior. It has been shown that consumption of alcohol to a rate of 0.05% of Breath Alcohol Content (BAC) increases the variance in trajectory by 3.2 on average [11], [3].[1] For the experiments, different scenarios of car flows are generated representing typical car trajectories; e.g. roundabouts, traffic lights, etc. An example of sequence used for the simulation is displayed in Fig. 2. The set of data is divided into 3 subsets of 11,900 samples (feature vectors), each representing 50 tracks of $N_S = 238$ steps. Two sets of trajectories with variance equal to 2 are used to train and test the system on normal behavior. The third set with variance equal to 6.4 is used to test abnormal behavior.

The algorithm described in this paper is tested on the aforementioned datasets. A truncated version $\hat{\xi}$ of the Gaussian ξ is implemented in the algorithm to reduce the computational load. The Gaussian truncated kernel $\hat{\xi}(\delta_{\mathbf{S}})$ is defined as

$$\hat{\xi}(\delta_{\mathbf{S}}) = \begin{cases} \xi(\delta_{\mathbf{S}}) & \text{if } \xi(\delta_{\mathbf{S}}) > \lambda\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{18}$$

The threshold T that defines the boundary between normal and abnormal behavior sets the sensitivity of the system. However, if the distribution of φ is assumed multi-Gaussian and if the centers μ_l have converged to the true means, the value of T can be determined during the learning stage as $T = \eta \max\left(\mathrm{var}_k(\varphi)\right)$. The value of η depends on the prior probability (ρ) of abnormal behavior and is defined as $\eta = \sqrt{2}\,\mathrm{erf}^{-1}(\rho)$, where $\mathrm{erf}(.)$ is the well-known Gaussian error function. This value determines the upper bound of T. The parameter λ is set to 0.001 and the standard deviation σ to 1.8. In the experiments, the counter n_l is initialized to 1 and the variance of the normal distribution r is set to 0.8. The cluster centers are initialized with random values to span the feature space.

B. Local vs. Global Distribution Learning

In this section, the performance of the local stochastic learning is compared to that of its global counterpart. Here, the global stochastic learning consists of a set of 714 clusters distributed over the entire feature space. The feature vector $\mathbf{X}$, for the global approach, is thus composed of both local and behavioral components:

$$\mathbf{X} = \begin{pmatrix} \mathbf{S} \\ \varphi \end{pmatrix} = \begin{pmatrix} x \\ y \\ x_t - x_{t-1} \\ y_t - y_{t-1} \end{pmatrix} \tag{19}$$

The local approach is implemented with the proposed algorithm, i.e. the spatial component $\mathbf{S}$ is used to index the local

[1] A BAC of 0.05% is the standard limit in Australia and most European countries.

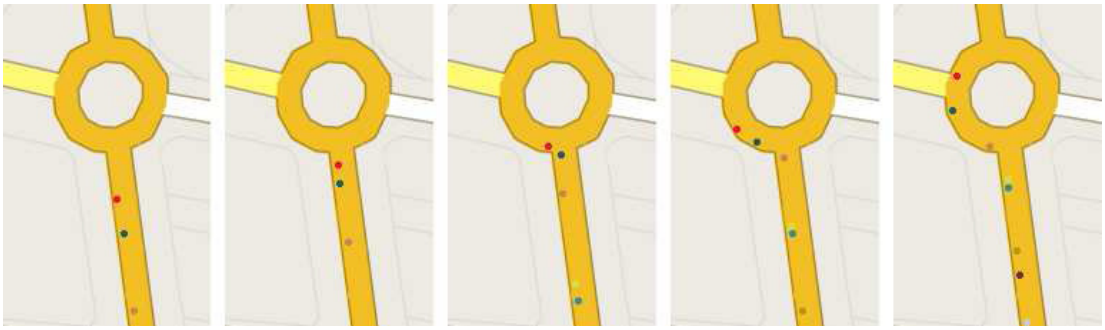


Fig. 2. Example of sequence representing the tracks used in the experiments. The synthetic sequences are generated on road maps in order to provide realistic scenarios; e.g. intersections, roundabouts, turns, etc.. The dots represent cars on the road. From *left* to *right*: frames 10, 14, 18, 22 and 28.

probability density of the behavioral component φ. Figure 3 is the ROC curve for both implementations. It is clear that the local approach performs better for the entire range of false detection. For instance, a false detection rate of 10% leads to a correct detection rate of 39% for the global approach whereas a 89% rate is achieved with the local approach. The latter performs better because the dimensionality of the feature vector is reduced and because there is no error on the position of the object since it is not estimated. For the global approach, error is introduced by the marginal local distribution estimation. Indeed, for a given number of clusters, the global approach shows a larger average distance between the clusters center and the feature vectors than the local approach. The global approach fails to reach a correct detection rate of 90% for a false detection rate of 40%. The inadequate estimation of local features distribution is mostly accountable for such a low performance.

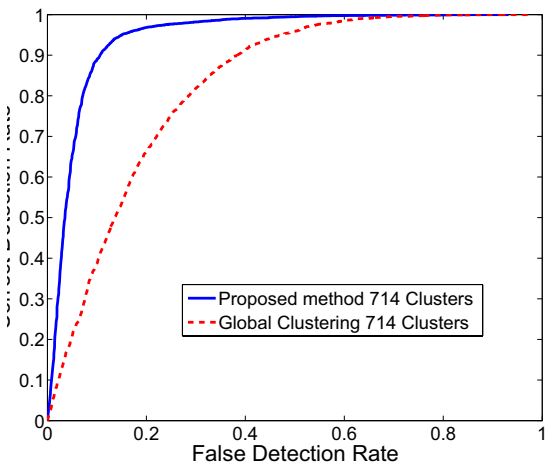


Fig. 3. ROC curves of stochastic learning algorithm for abnormal behavior detection. Comparison between the local implementation and the global implementation.

C. Local Distribution Learning vs. Self Organizing Maps

The proposed algorithm is compared with the Self Organizing Map (SOM) developed by Dahmane and Meunier [5]. For comparison purposes with [5], the behavioral component of the feature vector is again limited to the vector flow:

$$\varphi = \begin{pmatrix} x_t - x_{t-1} \\ y_t - y_{t-1} \end{pmatrix}, \tag{20}$$

where x and y are the cartesian coordinates of the object position. Note that this component can be extended to other features such as object acceleration, shape, size, etc. (e.g. [13]). The feature vector for the SOM is given by Eq. (19). SOMs have proven to give good results on abnormal behavior detection because of their property of topology conservation [5], [13]. This characteristic is particularly desirable when the feature vector is based on position since the neighborhood of the winning neuron is updated with the feature vector. The inclusion of the neighborhood confers the proposed approach with the topology advantage of SOM approach, whilst decreasing the feature vector size by the dimensionality of the spatial coordinates.

The performance of the proposed algorithm is compared to that of a SOM which models the global probability distribution. The proposed approach models the local probability distribution with a fixed number of clusters K; thus, the total number of clusters required is $K \times N_S$. For the SOM, the number of neurons is $h \times w$ where h and w are the width and the height of the map. We compare the SOM used by Dahmane and Meunier [5] to model the punctual context and the proposed algorithm. For comparison purposes, the map of the SOM is composed of 729 neurons (size $[27 \times 27]$) and the proposed algorithm is trained with 714 clusters, $K = 3$ for each object position. Figure 4 displays the ROC curves of both algorithms. It can be inferred that the proposed algorithm gives better performance (higher detection rate for a fixed false alarm rate) for correct detection rates above 60%. Note that a high rate of correct detection takes precedence over low false

TABLE I
CORRECT DETECTION RATE VERSUS NUMBER OF CLUSTERS.

False Detection	7.5%	10%	12.5%	15%	17.5%	20%
238 clusters	76.8%	83.5%	86.0%	88.0%	89.2%	90.4%
476 clusters	77.7%	85.6%	89.7%	92.0%	93.6%	94.9%
714 clusters	81.0%	89.1%	92.9%	94.7%	95.6%	96.5%
952 clusters	81.7%	89.9%	94.0%	96.2%	97.1%	97.5%
1190 clusters	82.7%	90.2%	95.5%	98.0%	99.0%	99.3%

TABLE II
CORRECT DETECTION RATE VERSUS SIZE OF SOM.

False Detection	7.5%	10%	12.5%	15%	17.5%	20%
Size [15 16] (240 Clusters)	50.5%	55.6%	61.1%	64.7%	68.2%	70.6%
Size [22 22] (484 Clusters)	63.3%	67.7%	71.6%	74.4%	77.1%	79.4%
Size [27 27] (729 Clusters)	64.6%	69.9%	73.9%	76.7%	79.6%	82.0%
Size [31 31] (961 Clusters)	81.5%	85.0%	87.6%	89.3%	90.7%	91.8%
Size [34 35] (1190 Clusters)	85.8%	89.6%	91.7%	93.2%	94.3%	95.0%

detection in most applications. The SOM and the proposed method have also been compared for different number of clusters/neurons. The results are presented in Tables I and II. The detection rate increases with the number of clusters/neurons for a given false detection rate in both cases. However, the local approach systematically outperforms the SOM, except for a false detection rate of 7.5% with 1190 clusters.

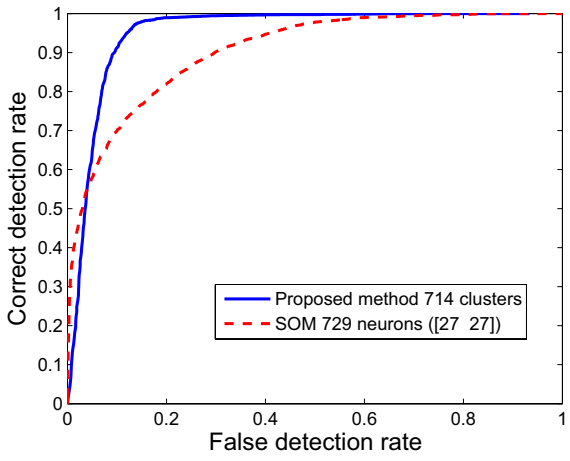


Fig. 4. ROC curves for the proposed technique and the SOM developed by Dahmane and Meunier [5].

VI. CONCLUSION

This paper presented a new technique for Abnormal Behavior Detection, based on local distribution of the feature vector. The motivation underlying such a local representation is the sparsity of the spatial component in the feature space. The integration of the neighborhood information into the estimation of the probability distribution reduces the dimensionality of the feature vector since the position information is not included explicitly in the feature vector. A stochastic clustering algorithm is used to learn the local probability distributions. The distribution is updated with incoming events, following a stochastic procedure, ensuring the clusters are optimally spread. Experimental results show that higher detection rates are achieved with fewer clusters compared to a global clustering approach.

REFERENCES

[1] A. Bouzerdoum. A stochastic competitive learning algorithm. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2, pages 908–913, 2001.

[2] H. Breit and G. Rigoll. A flexible multimodal object tracking system. In *Proceedings of IEEE International Conference on Image Processing*, volume 3, pages 133–136, 2003.

[3] B. Carswell and V. Chandran. Automated recognition of drunk driving on highways from video sequences. In *Proceedings of IEEE International Conference on Image Processing*, volume 2, pages 306–310, 1994.

[4] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[5] M. Dahmane and J. Meunier. Real-time video surveillance with self-organizing maps. In *Proceedings of the Canadian Conference on Computer and Robot Vision*, pages 136–143, 2005.

[6] Z. Dong, D. Gatica-Perez, S. Bengio, and I. A. McCowan. Semi-supervised adapted hmms for unusual event detection. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 611–618, 2005.

[7] B. Han, Z. Ying, D. Comaniciu, and L. Davis. Kernel-based bayesian filtering for object tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 227–234, 2005.

[8] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition. *Image and Vision Computing*, 14(8):609–615, 1996.

[9] N. Johnson and D. Hogg. Representation and synthesis of behaviour using gaussian mixtures. *Image and Vision Computing*, 20(12):889–894, 2002.

[10] C. Kwok, D. Fox, and M. Meila. Real-time particle filters. *Proceedings of IEEE*, 92(3):469–484, 2004.

[11] Y. C. Liu and C. H. Ho. The effects of different breath alcohol concentration and post alcohol upon driver's driving performance. In *Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management*, pages 505–509, 2007.

[12] S. Rao and P. S. Sastry. Abnormal activity detection in video sequences using learnt probability densities. In *Proceedings of Conference on Convergent Technologies for Asia-Pacific Region*, volume 1, pages 369–372, 2003.

[13] L. Xiu-Xiu, Z. Yan-Ning, and A. He-Jin Yuan. Activity analysis based on som. In *Proceedings of the International Conference on Machine Learning and Cybernetics*, volume 7, pages 3975–3979, 2007.