Detection of Abnormal Behaviour in a Surveillance Environment Using Control Charts

Stefan Hommes, Radu State, Andreas Zinnen, Thomas Engel
University of Luxembourg, SnT
6, rue R. Coudenhove-Kalergi, L-1359 Luxembourg
{stefan.hommes, radu.state, andreas.zinnen, thomas.engel}@uni.lu

Abstract

This paper introduces a new approach to unsupervised detection of abnormal sequences of images in video surveillance data. We leverage an online object detection method and statistical process control techniques in order to identify suspicious sequences of events. Our method assumes a training phase in which the spatial distribution of objects is learned, followed by a chart-based tracking process. We evaluate the performance of our method on a standard dataset and have implemented a publicly available open-source prototype.

1. Introduction

The security of public scenes or buildings is assured by surveillance officers and many cameras are deployed to observe various parts of a scene. As such this systems rely on a human operator, who has to rate the actions of many scenes in parallel; boredom and performance appraisal of the operator are common consequences. In a case of a crime, reaction is not fast enough and often the video recordings can only be used afterwards for forensic investigations. Automatic video surveillance is an approach that leverages machine learning by retrieving information and evaluating scenes as normal, unusual and abnormal without the help of a human operator.

A common approach to finding anomalies uses signature-based techniques. A drawback of these systems is that all possible actions have to be known and defined in advance and novel behaviours cannot be detected. Anomaly detection releases this constraint since no definition is needed, so decreasing the time taken in setting up the system. Furthermore, the main advantage is that patterns of unknown anomalous situations can be detected without defining them in advance.

In many scenarios, certain areas should never be occupied by humans in the course of normal business. An example is the tracks of a train station were a person should never be, as this can be seen to be a dangerous situation; or people walking around on the main runway of an airport. To detect such occurrences automatically, our system is tracking all objects (events) and classifies new objects from their position to alert the operator. Our contribution is using a Gaussian Mixture Model for modelling the scene, and control charts for notification. Control charts are practical tools for performing statistical process control [10] that have been shown to be very effective.

Figure 1. People located at different positions in the scene

For the experimental setup, we used a dataset and consider the case where people moving on the street is normal, but people on the lawn is abnormal, as shown in figure 1. We aim to build a system that can learn, unsupervised, how to differentiate between normal and abnormal cases.

Therefore, we assume as system requirements:

- unsupervised learning and self-tuning of parameters
- on-line operating mode
- expandable to multi camera networks and scalable object tracking

Our contribution comprises the following key points:

1. An image-scoring method that is used by a process control technique in order to identify sequences of unusual events.
2. Quantitative analysis with respect to the probability of false positives and false negatives.
3. A developed prototype available under an open-source licence (datalink is not presented in the submission in order to comply with the double-blind submission requirements)

This paper is organized as follows: In section 2, we present related work. Our concept of object tracking and modeling the scene is presented in section 3 and anomaly detection with control charts in section 4. The evaluation and results are presented in section 5. Finally, we conclude in section 6 and discuss our further ideas.

2. Related Work

For detecting unusual events from video data, Adam *et al*. [1] defined a set of requirements and developed a system where multiple low-level monitors extract features from the video stream. Afterwards, the likelihood for every new event is be computed and compared with those in the buffer. A policy-driven approach is presented by Kreibich *et al*. [6] this separates the event from the upper layer that formulates policy. This model can also exchange events with other peers to allow distributed detection.
Another approach is the definition of composite events that describe certain actions in the scene [8] [13] [3]. Subsequently, low-level algorithms are used to create simple events that are compared with already-known patterns of anomalies. Unsupervised learning to discover motions [12] [11] with trajectories analyses sequences of events and searches for interesting activities.
For intrusion detection, Ye *et al*. [16] described a method that monitors changes in event intensity and uses probabilistic techniques to evaluate events in network traffic data. He also used Exponentially Weighted Moving Average (EWMA) control charts to search for anomalous changes in event intensity of both correlated and uncorrelated data. Control charts analysis to classify the posture of a person in image sequences where used by [4] *et al*.
Nair *et al*. [9] use Hidden Markov Models (HMM) to recognize anomalous behaviour of people in an office corridor. The HMM is used to find actions in a time sequence of moving objects and classifies them as normal, or if they do not fit the activity pattern anomalous . A drawback of this system is that all actions that are considered as anomalous must be represented as a HMM. Our approach relies on developing a method that does not require a priori knowledge and can self-tune and self-calibrate.

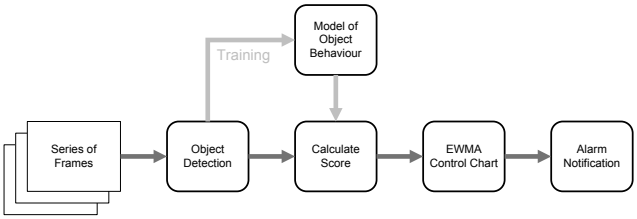

Figure 2. Architecture of an automated video surveillance solution

3. Tracking of Objects in the Scene

Our approach to an automated video surveillance solutions is presented in figure 2. The input can be images from a network video camera for on-line detection, or off-line from an archive. The latter is important for forensic analysis after a crime or for inspection of already existing video data.
After applying an algorithm for object detection to the image sequences, we record the object position during a training period and calculate a model for the activity in the scene. In operating mode, our model calculates a score to evaluate every new image which can be classified with control charts to alert the security officer if necessary.

3.1. Object Detection and Event Generation

The detection of people is an active research topic in image processing. Current algorithms perform well in a scene with a sparse density of people, but a crowd is hard to track due to occlusion. Paul Viola and Michael Jones [15] proposed an algorithm using Haar-like features for object detection. A classifier is trained with sample views of a particular object and the existence of oriented contrasts in the image can be used to find the object.
Each detected object is considered as an event e_i at a higher level of abstraction. An event is parameterized by the timestamp, the coordinates of the bounding box $(x, y, width, height)$, the object type and a reference to the image number in the sequence. The resolution of the object coordinates are equal to the size of pixels in width and height. We consider a set of events to be $E = \{e_1, ..., e_n\}$.

3.2. Modelling the probabilities using Gaussian Mixture Models

The probability of each object in the picture is determined by its position. Areas that are frequently covered by objects result in a higher object probability. To incorporate the probability of specific neighbours, we use a Gaussian Mixture Model (GMM) for continuous density estimation. The mixture model is a convex combination of Gaussian density functions (see figure[1] 3), where the probability of

[1] http://www.mathworks.com

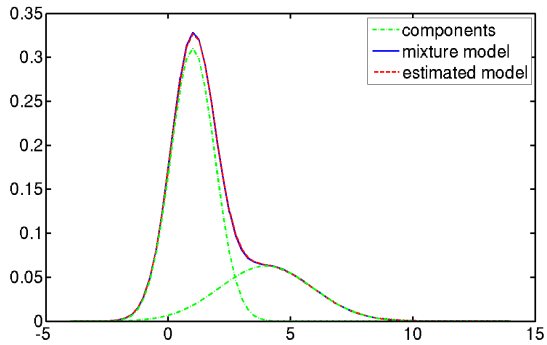

Figure 3. Example of Gaussian Mixture Model

an event can be described by the equation,

$$p(e_i) = w_1 N(e_1|\mu_1, \Sigma_1) + w_2 N(e_2|\mu_2, \Sigma_2) + \\ ... + w_n N(e_n|\mu_n, \Sigma_n) \quad (1)$$

where w_i are mixture weights that satisfy the constraints $\sum_{i=1}^{n} w_i = 1$ and $N(e_i|\mu_i, \Sigma_i)$ are the component Gaussian densities with mean vector μ_i and covariance matrix Σ_i. To determine the optimal number of Gaussians k for modeling the distribution, we calculate the total probability $P(E|M_k) = \sum_{i=1}^{n}(log(p(e_i)|M_k))$ for different k and choose the model M where the total probability is maximized.

4. Anomaly Detection in a Sequence of Images

In operating mode, the probability $p(e_i)$ for every new object can be calculated from the GMM. To evaluate a sequence of images, we first have to describe the anomaly from a single image that consists of an unknown number of events. Therefore, we define the probability of the image by the smallest event probability: $p(I_i) = \min(p(e_{(1)}, ...p(e_{(n)}))$.
The probability of a single image is not very robust for describing an anomalous situation in the scene. False positives from the object detector or sparse areas with fewer objects can result in low event probability for the image. To achieve a certain degree of robustness, we consider a sequence of anomalous images as a valid indicator. The evaluation of image sequences brings new information, since most actions are visible over several frames. Furthermore, it prevents low image probabilities from false positives from being considered as an alarm.

4.1. EWMA Control Charts

To classify a sequence of correlated images we propose the use of an Exponentially Weighted Moving Average (EWMA) chart [10]. This control chart type is used to monitor a process in real time, where current measurements have a lower impact then measurements that are further removed in time. This property can be used to decrease the weight of outliers, whereas an upper and lower control limit is used to classify new observations. In our scenario, we consider the image probability $p(I)$ as a measurement that has to be classified in normal or abnormal. The most recent EWMA data point can be calculated as:

$$EWMA_t = \lambda p(I)_t + (1 - \lambda) EWMA_{t-1} \quad (2)$$

where λ defines the impact of older data compared to new data. The mean of historical data $EWMA_0$, the upper control limit (UCL) and a lower control limit (LCL) are adapted with time and are defined by eq. 3 where $s^2_{ewma} = (\lambda/(2-\lambda)s^2)$ and s is the standard deviation.

$$UCL = EWMA_0 + ks_{ewma} \\ LCL = EWMA_0 - ks_{ewma} \quad (3)$$

The default value of factor k is 3 or chosen using the Lucas and Saccucci [7] tables and has to be adapted to the process data and behaviour.

4.2. Control rules

If $EWMA_t$ exceeds the upper or lower control limits, then something unusual has happened; otherwise the process is in control. In addition, several rules to evaluate the process can be applied if a value is out of the limits; these are defined as Western Electric Company Rules (WECO) [5] which represent the standard reference rules used in statistical process control:

Rule	Description
1	Any point above or below $k\sigma$
2	2 out of the last 3 points above or below $(k-1)\sigma$
3	4 out of the last 5 points above or below $(k-2)\sigma$
4	8 consecutive points above or below mean

As we are considering EWMA control charts, the threshold σ of UCL and LCL is set to s_{ewma}.

5. Experimental Results

The proposed methods were tested in a typical scenario that is representative. We consider a scene with a street where pedestrians are allowed to walk as a normal behaviour, but people that walk on the lawn are treated as abnormal.

5.1. Dataset

We used the benchmark data from the PETS 2009 workshop[2] for the experiment. This dataset consists of different

[2] http://www.cvg.rdg.ac.uk/PETS2009/a.html

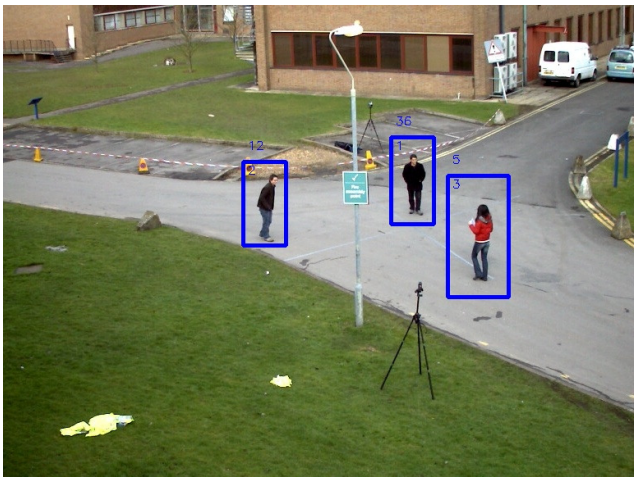

Figure 4. Three correctly classified objects in PETS 2009 dataset

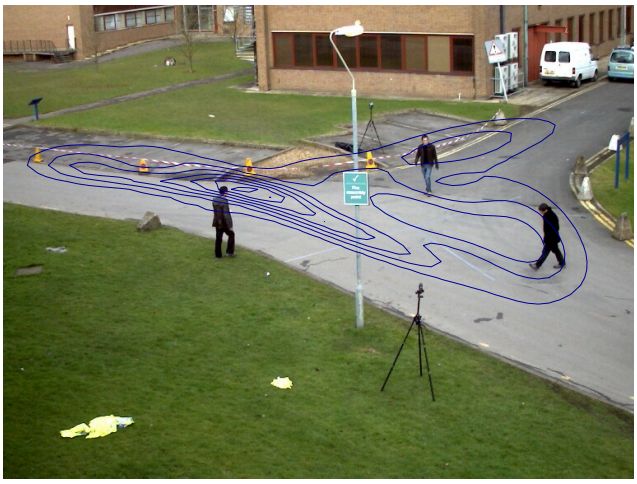

Figure 5. Contour plot of the GMM in the scene

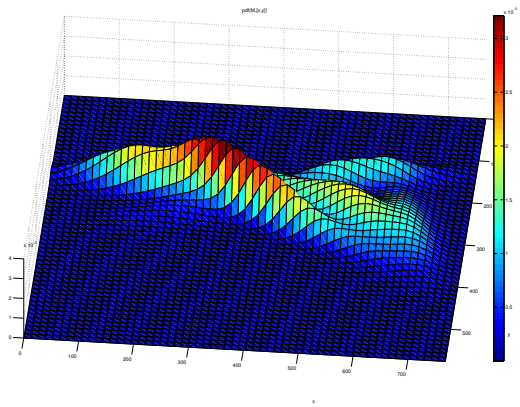

Figure 6. Density distribution of the GMM in the scene of figure 5

scenes, from which the camera view001 is chosen because it presents the best perspective for analysing people both on the street and on the lawn. The crowd activities vary from sparse to medium and high density crowds and show different kinds of behaviour (e.g. walking, running and loitering). The dataset consists of 4477 images and includes all images from the sub-datasets S1 to S3. All images have a resolution of 768x576 pixels and are compressed as JPEG image sequences.

5.2. Implementation

Object detection was done using the OpenCV [2] Python module with the algorithm from section 3.1. The experiments were run on a dual-core Intel PC with 2.80 GHz and 4 GB of RAM. The object detection algorithm was the most time consuming part and the complete calculations allowed a frame rate of 5 fps in on-line mode. A processed image with three correctly-classified objects is shown in figure 4.

The total number of detected events in the dataset from all images was 11832. Since the probability for every object has to be calculated, we use a GMM for density estimation as discussed in section 3.2. The number of Gaussians must be estimated and we chose the best GMM with 4 - 8 components. To determine the number of Gaussians for the dataset, we used a 10-fold cross-validation and calculated the total probability for every subset. We consider the optimal model is that where total probability is maximized. The contour plot of the GMM for the PETS 2009 dataset is presented in figure 5 and the 3D surface plot in figure 6.

To evaluate the classification of every event or image on the basis of the probability, we use receiver operating characteristic (ROC) curves. A ROC curve is a graphical plot for a classifier system in which the discrimination threshold is varied. The x-axis depicts the 1-precision or 1-specificity, and the y-axis the recall or sensitivity. The recall is defined as $\frac{tp}{tp+fn}$ and the precision as $\frac{tp}{tp+fp}$. The probability is used for the discrimination threshold in the ROC curve and varies from the lowest probability of the dataset to the highest.

The correct threshold depends on how often the security officer wants to be notified. If the threshold is set to a high probability, more people on the grass will be detected (true positives) and the recall increases. On the other hand, more people that were on the street are classified as people on the lawn, leading to a lower precision.

5.3. Results

To evaluate the results of correctly classified objects, we annotated an image with the labelMe tool [14]. We separated the image into two regions, considering the street and the footpath in the upper left corner as normal locations.

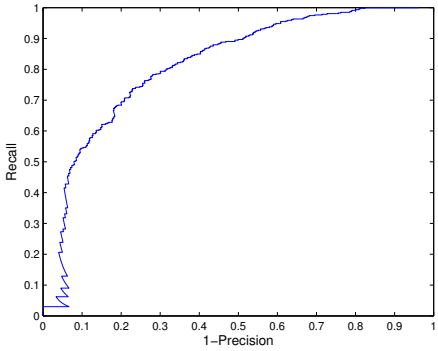

Figure 7. ROC curve of all events with probability as a threshold

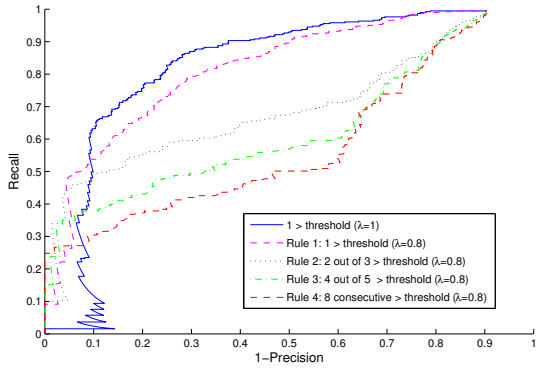

Figure 8. ROC curve of images with probability as threshold and WECO rules

5.3.1 Classification based on events

The receiver operating characteristic (ROC) curve is presented in figure 7. The parameter for the curve is the threshold of the event probability, which is smallest at the start point and increases to the highest probability. For every threshold the *recall* gives the percentage of objects that where found, compared to the overall quantity of objects on the lawn. The *1-precision* rate displays the percentage of objects that were classified as objects on the street. For a recall of 70% the results are very promising, since the precision is 80% and only 20% of objects on the street are classified as objects on the lawn.
Low probabilities from objects on the street cause some false positives. Because the training material did not include people in every possible normal position, the modelling of the scene after training had low object probabilities at certain positions on the street. For the same reason, the footpath in the left upper part of the image is not modelled well by the GMM due to the low object frequency in this area (see figure 6).

5.3.2 Classification based on images with control charts

The receiver operating characteristic (ROC) curve for the classification based on image probabilities and control charts is presented in figure 8. The solid line describes the classification without smoothing where $\lambda = 1$. The rules of the control charts can be used to identify different behaviours in the scene and notify the operator only under special conditions. If the WECO rule 1 is applied, smoothing from the EWMA control charts improves the classification for low probabilities in the ROC curve, but brings a small decrease in the precision rate. For example, if we consider rule 1 from the ROC curve in figure 8 and want to detect 70% of all images that contain people on the lawn from the dataset, we obtain a precision of 80%.

Rule 2 and 3 result in a even more stable outset and can be used if the operator should not be notified by false alarms and a lower detection rate is acceptable. Rule 4 can be useful to search for groups that gather on the lawn and ensure a low probability for a sequence of 8 consecutive images. A typical scenario is an emergency when all the people are trying to escape from a fire, or when there is a panic for some other reason.
Figure 9 shows the EWMA control chart for all image probabilities in the dataset, although only the lower control limit (LCL) is useful in our scenario and shows a correctly classified image with a person on the lawn. The correct UCL threshold depends on how often the security officer wants to be notified and can be set by the ROC curve in figure 8.

6. Conclusion and Future Work

In this paper we have presented a novel approach to video event analysis using control charts in a surveillance environment. Reducing the manual effort of analysing the information stream is mandatory in a multi-camera network. Our method achieves this goal, ensuring that only a small portion of video data is presented to the security officer. In the scenario described, the detection of people on the lawn is sufficient to deter most intruders. On the other hand, the security officer is not irritated by a huge number of false positives.
Although the detection of objects works encouraging with a low density of people, occlusion during crowd movements decreases the detection rate and is a drawback. We consider the use of a multi-camera network with a variety of viewpoints as an improvement to achieve a better object detection rate. We are also investigating temporal evaluation of objects and multivariate (EWMA) control charts.
A prototype of the proposed architecture has been built in Python and the GUI is displayed in figure 10.

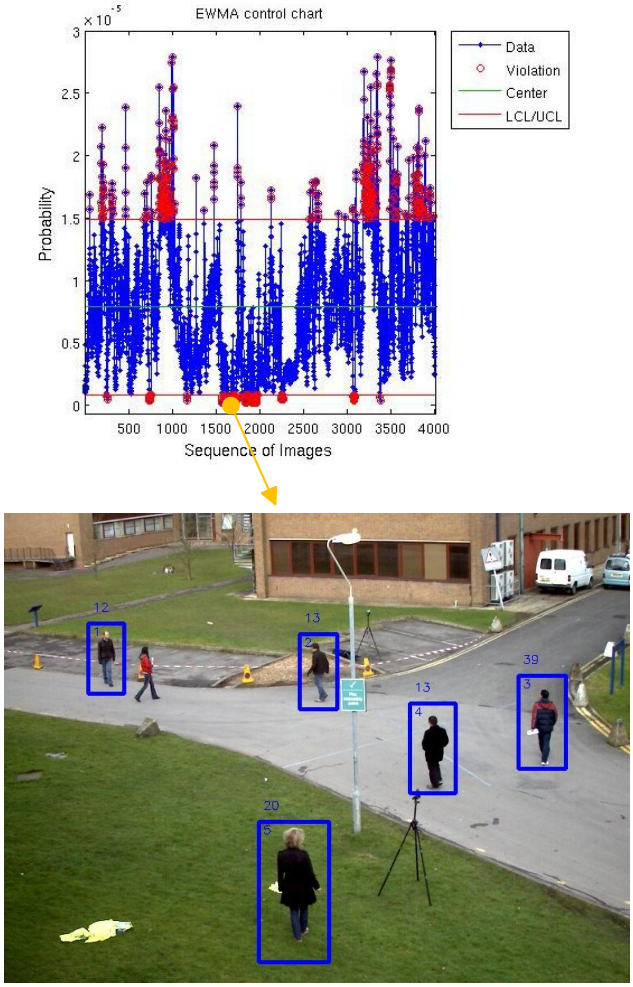

Figure 9. EWMA chart with the probabilities of all images in the dataset and $4s_{ewma}$ for UCL and LCL. A sample image with a probability below LCL and a person standing on the lawn is displayed below the chart.

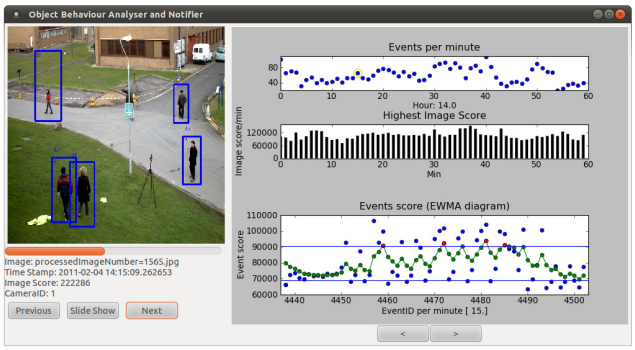

Figure 10. Prototype with Graphical User Interface (GUI) of our proposed surveillance architecture

Acknowledgement

The present project is supported by the National Research Fund, Luxembourg.

References

[1] 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*, 30:555–560, 2008.

[2] G. Bradski and A. Kaehler. *Learning OpenCV*. O'Reilly Media Inc., 2008.

[3] C. Brax, L. Niklasson, and M. Smedberg. Finding behavioural anomalies in public areas using video surveillance data. In *Information Fusion, 2008 11th International Conference on*, 30 2008.

[4] E. Elbasi, L. Zuo, K. Mehrotra, C. Mohan, and P. Varshney. Control charts approach for scenario recognition in video sequences. 2008.

[5] W. Electric. *Statistical Quality Control Handbook*. Western Electric Corporation, Indianapolis, Ind., 1956.

[6] C. Kreibich and R. Sommer. Policy-controlled event management for distributed intrusion detection. In *Distributed Computing Systems Workshops, 2005. 25th IEEE International Conference on*, pages 385 – 391, 6-10 2005.

[7] J. M. Lucas and M. S. Saccucci. Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics*, 32(1):pp. 1–12, 1990.

[8] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia. Event detection and analysis from video streams. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(8):873 –889, aug 2001.

[9] V. Nair and J. J. Clark. Automated visual surveillance using hidden markov models. In *Vision Interface*, page 88, 2002.

[10] J. S. Oakland. *Statistical Process Control*. Butterworth-Heinemann, 2008.

[11] L. Patino, F. Bremond, M. Evans, A. Shahrokni, and J. Ferryman. Video activity extraction and reporting with incremental unsupervised learning. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, 29 2010.

[12] G. Pusiol, F. Bremond, and M. Thonnat. Trajectory based activity discovery. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, 29 2010.

[13] R. ROMDHANE, F. BREMOND, and M. THONNAT. A framework dealing with uncertainty for complex event recognition. In *7th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS*, Aug 2010.

[14] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database and web-based tool for image annotation. *MIT AI Lab Memo*, 2005.

[15] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:511, 2001.

[16] N. Ye, S. Emran, Q. Chen, and S. Vilbert. Multivariate statistical analysis of audit trails for host-based intrusion detection. *Computers, IEEE Transactions on*, 51(7):810 –820, jul 2002.