HMM-based Abnormal Behaviour Detection Using Heterogeneous Sensor Network

Hadi Aliakbarpour[1], Kamrad Khoshhal[1], João Quintas[1],
Kamel Mekhnacha[2], Julien Ros[2], Maria Andersson[3], and Jorge Dias[1]
[1]ISR, University of Coimbra, Portugal; {hadi,kamrad,jquintas, jorge}@isr.uc.pt
[2]Probayes, France; {kamel.mekhnacha, julien.ros}@probayes.com
[3]FOI, Linköping, Sweden; maria.andersson@foi.se

Abstract. This paper proposes a HMM-based approach for detecting abnormal situations in some simulated ATM (Automated Teller Machine) scenarios, by using a network of heterogeneous sensors. The applied sensor network comprises of cameras and microphone arrays. The idea is to use such a sensor network in order to detect the normality or abnormality of the scenes in terms of whether a robbery is happening or not. The normal or abnormal event detection is performed in two stages. Firstly, a set of low-level-features (LLFs) is obtained by applying three different classifiers (what are called here as low-level classifiers) in parallel on the input data. The low-level classifiers are namely Laban Movement Analysis (LMA), crowd and audio analysis. Then the obtained LLFs are fed to a concurrent Hidden Markov Model in order to classify the state of the system (what is called here as high-level classification). The attained experimental results validate the applicability and effectiveness of the using heterogeneous sensor network to detect abnormal events in the security applications.

Keywords: Heterogeneous **s**ensor network, LLF (Low level Feature), HBA (Human Behaviour Analysis), HMM (Hidden Markov Model), LMA (Laban Movement Analysis), Crowd analysis, ATM (Automated Teller Machine) security.

1 Introduction

Recently, the demand for using automatic surveillance systems has been increasing. Many research areas, such as computer vision, signal processing, voice analysis and sensor fusion and pattern recognition are involved in this type of applications. Work on detection and tracking algorithms for dense crowds can be found in the literature. In [1] a method is suggested for simultaneously tracking all people in a dense crowd using a set of cameras with overlapping fields of view. In [2] a real-time system for detection of moving crowds is presented. HMM has been used in various applications for behavior recognition, e.g. [3] for facial action recognition and [4] for crowd behavior analysis. Crowd analysis can be used to get an understanding of the crowd as a whole, without any detailed information on individuals in the crowd. Crowd activity provides information which can be used to detect if there are people running or fighting [4]. A deep contribution in the field of human-machine interaction (HMI), based on the concept of LMA (Laban Movement Analysis), is presented by Rett and Dias in [5]. In their work a Bayesian model is used for learning and classification. The LMA is

presented as a concept to identify useful features of human movements to classify human gestures. Frequency-based extracted features are used in a LMA-based approach in our previous work for the sake of behaviour analysis [6]. Using audio signals for security purposes is proposed in [7]. Using distributed sensor network for the surveillance is investigated and proposed by Aliakbarpour et al. in [17,18].

A two-staged classification approach, to detect abnormal events in a security scenario, is introduced in this paper. Firstly, a set of low-level-features (LLFs) is obtained by concurrently applying three different classifiers (what are called here as low-level classifiers) on the input data. The low-level classifiers are namely LMA, crowd and audio analysis. Then the obtained LLFs are fed to a concurrent HMM in order to classify the state of the system (what is called here as high-level classification). A network of heterogeneous sensors such cameras and microphone arrays are used in a synergic way to observe the scene.

This paper is arranged as following: Our contribution to sustainability is presented in Sec. 2. Low-level classification is introduced in Sec. 3. A HMM-based classification (here is known as the high level classifier) is discussed in Sec. 4. Sec. 5 is dedicated to the experimental results and eventually the conclusion is presented in Sec. 6.

2 Contribution to sustainability

In order to protect citizens, property and infrastructure, surveillance systems are increasingly being used. Commonly, these surveillance systems are installed in public spaces, covering large areas, where a great number of people populate the camera's fields of view. Consequently such systems consist often of a large amount of distributed sensors, typically CCTV cameras, monitored by operators in a control room. Since humans possess a limited capacity of driving its focus of attention, it is impossible to the system's operators pay attention to all what is happening in all monitors at a given time. Moreover, to recognizing abnormal or threatening events is a complex cognitive task requiring a focus that humans can uphold for only a short time. Therefore, there is the need for a persistent system capable of making a pro-active surveillance. In this paper we introduce a method to detect abnormal situation, in the ATM scenarios, using the observations of a heterogeneous sensor network comprising of cameras and microphone arrays.

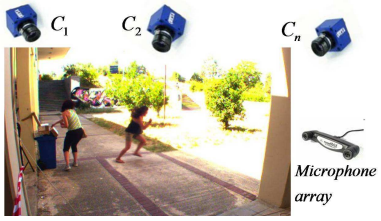

Figure 1- A superimposed view of the ATM scenario when a rubbery is happening.

3 Low-level classification

As mentioned, the idea is to apply a two staged procedure in order to conclude whether the scene's state is normal or abnormal. Here the used data is from an exclusive multimodal database, referred as PROMETHEUS database[1] [8]. This database is in support of the development and the evaluation of the algorithms which are intended to analyze and identify human actions and behaviors in the context of surveillance using multi-modal approaches. It comprises of various securities scenarios and among them

[1] www.prometheus-FP7.eu.

we focused on the ATM ones. Many typical events such as walking, waiting, taking money and some atypical events such as robbery have happened in the ATM scenarios. The intention is to automatically identify and detect the state of the scene in terms of whether a normal or abnormal event is happening.

The scenes are observed by a network of heterogeneous sensors, composed of video cameras, thermal cameras and microphone arrays. Among the heterogeneous data in the database, we selected two modalities: image and sound. Three different low-level classification methods, namely LMA, crowd and audio analysis, have been applied on these data in order to obtain a set of LLFs. Table 1 summarizes the inputs and outputs for each method. As can be seen, two of these methods, the LMA and crowd analysis, have their inputs from applying some preliminary data fusion and tracking algorithms (which are supposed to be available) on the raw data and just the audio analysis one has a direct input from the raw sound signals. For the sake of having a low level classification on the sound signals, the method introduced in [7] has been used. Here we continue to introduce the LMA and crowd analysis methods.

Table 1: Inputs and outputs for the different methods in the low-level classification stage**.**

Method	Input	Output
Crowd Analysis	Optical flow from image data	Crowd ratio
Laban Movement Analysis	3D positions of heads and feet (available from tracking algorithms)	Pr(walking), Pr(running), Pr(falling) and Pr(standing)
Audio Analysis	Sound signals	LL-ratio

3.1 LMA-based human movement classification

LMA is a well-known method to describe and analyze human movements by several components; Body, Space, Shape, Effort and Relationship [9]. Each component describes human movements by different aspects. Among the different components of LMA, here we have selected Effort component to observe human movements in terms of how motion of human body parts are happening with respect to inner intention (our recent work in [6]). Effort has four sub-components and each of them has two states; Effort.time (sudden/sustained), Effort.space (direct/indirect), Effort.weight (light/strong) and Effort.flow (bounded/free). In this work, the Effort.time component for two body parts (head and feet) is estimated. Here the interesting movements to classify are standing, walking, running and falling-down. A Bayesian Network (BN), which is a well-known "tool" to deal with uncertainty, has been used for our LMA parameters estimations (to classify different human movements). The BN has three levels;

- The lowest level comprises frequency-based features which were achieved from Power Spectrum (PS) technique that applied on acceleration signals of body parts. First four coefficients were collected from PS signal of each body part acceleration signals. Each coefficient has four state possibilities (see [6]) which are defined by several thresholds.
- Middle-level includes LMA parameters, Effort.time component of the body parts.
- The highest level is just one node which has the number of interest human movement's states.

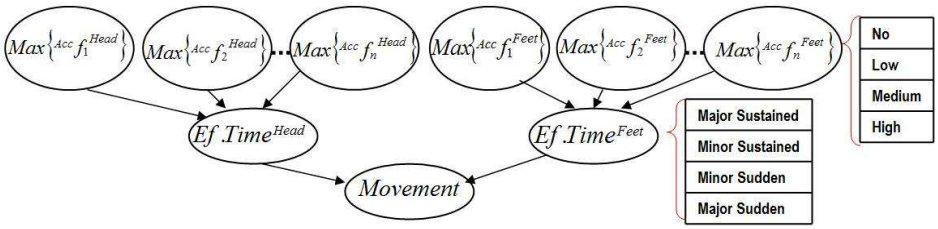

Figure 2. LMA-based Bayesian net

3.2 Crowd analysis

Crowd analysis is used to get an understanding of the crowd as a whole, without any detailed information on individuals in the crowd (see our recent work [4]). The aim is to get an approximate understanding of the activity level of the crowd as well as the size of the crowd. Crowd activity provides information which can be used to detect if there are people running or fighting. In general, normal crowd behavior often corresponds to calm movements, where people are standing or moving slowly through the scene, without making excessive gestures. Abnormal behaviour is instead likely to be accompanied by more rapid movements. The different levels of crowd activity are estimated by optical flow calculations, which estimate the relative motion between consecutive images. If a person is walking quickly, running, or moving his or her arms rapidly, the magnitude of optical flow will be larger compared to the case when a person is moving slowly or standing still. The different levels of crowd activity can be derived by using for example the following approaches:

1. Manually setting the different levels by observing the optical flow under known conditions, i.e. when the different types of events in the scene are known.
2. Applying a more automatic approach for obtaining the levels by using basic statistics (mean value and standard deviation) on relevant training data.

The crowd size provides information that can be used for getting warnings of forthcoming fights, attempted robbery and risk for riots. What is considered to be a large or small crowd will differ from case to case. For example, in a city area a large crowd may be 20-25 persons or more. At large sport events, large crowds are probably hundreds or thousands of persons. Fighting and robbery at the ATM means that at least two persons are present on a small area at the same time. The crowd size estimate is obtained by first performing background subtraction to obtain the foreground pixels. We assume prior knowledge of the approximate number of pixels per person, which depends on the distance between camera and crowd. The number of people is then obtained by dividing the total amount of foreground pixels by the number of assumed pixels per person. Also crowd growth rate and crowd density can be of interest to understand the crowd behaviour. Is the crowd growing quickly and/or are the people standing close to each other? Crowd growth rate can be estimated by studying the change in crowd size for a certain reasonable time period. Crowd density can be estimated by relating the crowd size to a specific area in the image. By combining estimates of crowd activity and crowd size, uncertainty in the classification of behavior will be reduced. For example, increased activity (a running person) at an ATM *together* with the information that there were at least two persons present close to the ATM, will strengthen the view that there have been an attempted robbery.

4 Concurrent HMM-based classification

For behaviour recognition, we are interested in detecting the current behavior amongst N known behaviors (i.e. the behaviour library). For this purpose, we propose to use a concurrent HMM architecture.

4.1 Principle

A concurrent hidden Markov model is composed of several hidden Markov Models, each one describing one class (see Fig.3-left). To summarize, the concurrent HMM centralizes the on-line update of the behaviour belief and contains:

1. The set of HMMs representing basic behaviour library (one HMM per behaviour);
2. The transition between behaviors model that could be either defined by hand (by an expert), or learnt from annotated data.

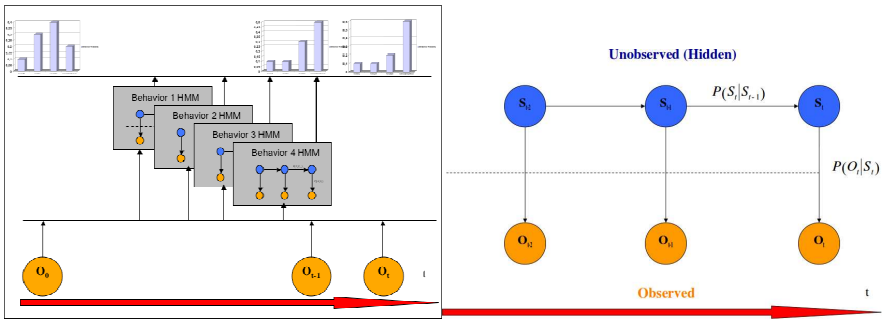

Figure 3: Concurrent Hidden Markov Models (left) HMM model (right)

Hidden Markov Models are used to characterize an underlying Markov chain which generates a sequence of states. The term Hidden in the HMM name comes from the fact that the sequence of states is not directly observable. Instead the states generate an observable sequence. Thus, the output depends on the current state and on previous outputs. These tools are widely used in the field of sound processing [11], gene finding and alignment in DNA sequences [12]. They were introduced by Andre Markov in [13] and developed in [14]

HMM are widely employed in the field of computer vision to recognize gesture or human behaviour [15]. In these applications, the observation variables are features extracted for video data. The principle of an HMM is presented on Figure Fig.3-(right) in which S represents the state variable and O represents the observation ones. In our application, each HMM describes a human behaviour and is learned using a training dataset composed of labeled observation sequences that are low-feature extracted from the different and thanks to three different approaches, namely Laban Movement Analysis (LMA), crowd and audio analysis.

4.2 Construction/Learning

Constructing a concurrent hidden Markov model consists in:

- Learning the set of HMM models representing the behaviour library (one HMM per behaviour) using an annotated data set.
- Defining the transition matrix between the behaviors. This transition model could be either defined by hand (by an expert), or learnt from an annotated data set.

Learning the behaviour transition model is straightforward and consists in computing simple statistics (histograms) of transitions using the annotated data set.

Learning the underlying HMM models (a HMM per behaviour) is more complex. It can be divided into two sub-problems:

1. Finding the optimal number of states *N*. The optimal number of internal states within the HMMs could be chosen by hand thanks to an expert. In this case no algorithm is needed and the learning of the HMM is reduced to the learning of its parameters. However, since an HMM is a Bayesian Network, a score that allows a compromise between fitting learning examples (D) and the ability of generalization (see the Occam Razor Principle) can be employed to find it automatically [15]. For example the classical Bayesian Information Criterion [16] that maximizes the likelihood of the data while penalizing large size model can be used:

$$BIC(n, D) = \log(likelihood(D, n)) - \frac{1}{2} \times nparams(n) \times \log(|D|)$$

In this case the optimal number of states is given by: $n* = \arg\max_n BIC(n, D)$

2. Learning the parameters of the HMM given N (i.e., the transition matrix $P(S_t \mid S_{t-1})$), the observation distribution $P(O_t \mid S_t)$, and the initial state distribution $P(S_0)$). The idea is find the parameters that maximize the data likelihood. For this purpose the methods generally employed are the classical EM algorithm (aka Baum-Welch algorithm in the HMM context), or the Iterative Viterbi algorithm.

4.3 Recognition

As previously emphasized, the concurrent hidden Markov model is used to recognize on-line or off-line the current behaviors amongst *N* known behaviors. This is easily performed by finding the HMM *M* that maximizes $P(M \mid O_{t-n},...,O_t)$ for the off-line case (or $P(M \mid O_t)$ for the on-line case).

5 Experiments

As mentioned, the ATM scenarios of PROMETHUS [8] multimodal database are used in this paper. There are four selected ATM scenes with different durations. The interesting event which we call as abnormal state in this kind of scenario is robbery. Based on the proposed approach the process of event detection has two levels for performing the classification. Firstly, some low-level features are obtained by using three different approaches (see Fig.3). Then these low-level features are fed to a HMM as a high level classifier in order to estimate the scene's state. Apart of these LLF inputs for the HMM, another parameter which is named as "*environment parameter*" is also consider for the HMM. The *environment parameter* is defined as the relative positions of people to the ATM. In the context of the ATM scenario in PROMETHEUS dataset, we defined the robbery state as when the robber waits in

ATM's area, approaches a person who is taking money from the machine, steals the money and then rapidly escapes.

In the database, there are 139 samples corresponding to the normal situations and 8 samples corresponding to the robbery (abnormal) situations. Each sample has a 10 seconds long. Among these samples, 61 samples of normal data and 4 samples of abnormal ones have been randomly selected for HMM learning process and the others (78 samples of normal and 4 samples of abnormal) for HMM classification process.

Here we have implemented different experiments on the data in order to demonstrate the applicability and effectiveness of using heterogeneous data fusion in the proposed manner. Table 2 depicts the result of the HMM-based classifier when the output of each low-level classifier is individually applied to the high-level (final) classifier. Then the experiment is performed when a pair of the low-level-classifier outputs is used for the classification (three possible pair combinations, see Table 3). Eventually all of the low-level classifier's outputs have been fed to the high-level classifier. Table 4 depicts the result of this last case, in which there are the best percentages of the true even detections. It validates the effectiveness of the proposed method for the sake of surveillance applications

Table 2- High-level classification result: when just one of the three low-level features has been used.

Method	LMA			Crowd			Sound(LL ratio)		
	Normal	Robbery	%	Normal	Robbery	%	Normal	Robbery	%
Normal	72	6	92	64	14	82	76	2	97
Robbery	0	4	100	1	3	75	1	3	75

Table 3- High-level classification result: Three possible combinations of using a pair of low-level features.

Methods	LMA + Crowd			LMA + Sound (LL ratio)			Sound (LL ratio) + Crowd		
	Normal	Robbery	%	Normal	Robbery	%	Normal	Robbery	%
Normal	72	6	92	77	1	98	77	1	98
Robbery	0	4	100	0	4	100	1	3	75

Table 4- High-level classification result: Fusion of all low-level features

Method	LMA + Sound (LL ratio) + Crowd		
	Normal	Robbery	%
Normal	77	1	98
Robbery	0	4	100

6 Conclusion

Abnormal human behavior detection by using a network of heterogeneous sensors, in some ATM scenarios, has been proposed in this paper. A two-staged classification, divided as low-level-classification and high-level-classification, is used in order to detect the abnormality of the current state of the scene. Here, the interesting abnormal event is defined as happening a robbery near the ATM. The LMA, crowd analysis and audio analysis are the methods which are used in the low-classification stage. For the sake of high-level classification, a concurrent HMM is applied. The attained experimental results validate both the applicability and efficiency of the proposed method for the sake of surveillance applications.

Acknowledgment: *Hadi Ali Akbarpour* is supported by the **FCT** (Portuguese Fundation for Science and Technology). This work is supported by the European Union within the FP7 Project

PROMETHEUS, www.prometheus-FP7.eu. The authors would like to thank our partners from University of Patras for providing sound data.

References

[1] Eshel, R. & Moses, Y. Homography Based Multiple Camera Detection and Tracking of People in a Dense Crowd. CVPR 2008. IEEE Conference on, 2008.

[2] Bird, N.; Atev, S.; Caramelli, N.; Martin, R.; Masoud, O.; Papanikolopoulos, N.; , "Real time, online detection of abandoned objects in public areas," Robotics and Automation, 2006. ICRA 2006. ,IEEE, vol., no., pp.3775-3780, 15-19 May 2006.

[3] Lifeng Shang; Kwok-Ping Chan; , "Nonparametric discriminant HMM and application to facial expression recognition,". CVPR 2009. IEEE, vol., no., pp.2090-2096, 20-25 June 2009.

[4] P. Drews, J. Quintas, J. Dias, M. Andersson, J. Nygards, J. Rydell - Crowd behavior analysis under cameras network fusion using probabilistic methods, the 13th International Conference on Information Fusion, 26-29 July 2010 EICC Edinburgh, UK.

[5] Joerg Rett, Jorge Dias, and Juan-Manuel Ahuactzin. Laban Movement Analysis using a Bayesian model and perspective projections. Brain, Vision and AI, 2008.

[6] Kamrad Khoshhal, Hadi Aliakbarpour, Joao Quintas, Paulo Drews, and Jorge Dias. Probabilistic LMA-based classification of human behaviour understanding using power spectrum technique. In 13th International Conference on Information Fusion2010, EICC Edinburgh, UK, July 2010.

[7] S. Ntalampiras, I. Potamitis, and N. Fakotakis, “An Adaptive Framework for Acoustic Monitoring of Potential Hazards”, EURASIP Journal on Audio, Speech, and Music Processing Volume 2009 (2009), doi:10.1155/2009/594103.

[8] Ntalampiras, S.; Ganchev, T.; Potamitis, I. & Fakotakis, N. Heterogeneous Sensor Database in Support of Human Behaviour Analysis in Unrestricted Environments: The Audio Part The seventh international conference on Language Resources and Evaluation (LREC), 2010.

[9] L. Zhao and N.I. Badler. Acquiring and validating motion qualities from live limb gestures. Graphical Models, pages 1_16, 2005.

[10] Guangyi Shi, Yuexian Zou, Yufeng Jin, Xiaole Cui, andWen J. Li. Towards HMM based human motion recognition using mems inertial sensors. In Proceedings of the 2008 IEEE, International Conference on Robotics and Biomimetics, 2009.

[11] Lawrence R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. pages 267–296, 1990.

[12] Pachter L., Alexandersson M., and Cawley S. Applications of generalized pair Hidden Markov Models to alignment and gene finding problems. In *RECOMB ’01: Proceedings of the fifth annual international conference on Computational biology*, pages 241–248, New York, NY, USA, 2001. ACM.

[13] Markov A. An example of statistical investigation of the text eugene onegin concerning the connection of samples in chains. In *Lecture at the physical-mathematical faculty, Royal Academy of Sciences, St. Petersburg*.

[14] Baum L. E., Petrie T., Soules G., and Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

[15]Biem A. A model selection criterion for classification: Application to HMM topology optimization. In *ICDAR ’03: Proceedings of the Seventh International Conference on Document Analysis and Recognition*, page 104, Washington, DC, USA, 2003.IEEE.

[16] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[17] Hadi Aliakbarpour, J. F. Ferreira, Kamrad Khoshhal and Jorge Dias . A Novel Framework for Data Registration and Data Fusion in Presence of Multi-modal Sensors, Proceedings of DoCEIS’10, IFIP AICT 314/2010, Springer, pp 308-315.

[18] Hadi Aliakbarpour and Jorge Dias. Human Silhouette Volume Reconstruction Using a Gravity-based Virtual Camera Network. In the Proceedings of the 13th International Conference on Information Fusion, 26-29 July 2010 EICC Edinburgh, UK.