USING CONCURRENT HIDDEN MARKOV MODELS TO ANALYSE HUMAN BEHAVIOURS IN A SMART HOME ENVIRONMENT

João Quintas; Kamrad Khoshhal*; Hadi Aliakbarpour*; Martin Hofmann**; Jorge Dias**

*Institute of Systems and Robotics, University of Coimbra
**Technische Universität München

ABSTRACT

This paper addresses learning and recognition of human behaviour models from multi-modal observations in a smart home environment. The proposed approach successfully implements concurrent Hidden Markov Models that identify the occurring situation. This approach corresponds to the high-level part from a framework to obtain high-level classification of human behaviour analysis. The results were obtained for a smart home environment, where cameras, microphones and a PMD sensor were deployed. The sensory information was first fed to the low-level classification stage, where it was analysed by four different classifiers which generate the observations to the high-level classification stage. For each situation an HMM is used, allowing the fusion of the data provided by the different sources present in the low-level classification stage. This approach proved to be highly scalable, since the recognition of new situations can be accomplished by means of adding the adequate HMM.

1. INTRODUCTION

As "smart" technologies evolve, the concept of smart environments or smart homes also changes. The concept that a smart home is capable of responding to elaborated commands and performs simple tasks is starting to be insufficient for taking into account the upcoming interactive technologies. There is an emerging trend that preconizes that smart homes should be auto-adaptable and self-reconfiguring habitats. This concept envisages environments where the needs, moods and lifestyles of its inhabitants are taken into consideration to provide better services and life conditions, leading to the pursuit for symbiotic and user-centred environments focused on the well being of people. As more sophisticated interaction paradigms are proposed, we will witness the implementation of increasingly complex sensory environments which, by combining different types of sensors, will try to communicate with the users using natural language. This will compel researchers to investigate new concepts, methods, algorithms, and implementations of multi-modal networks of heterogeneous sensors.

Smart homes will need, in some cases, to be capable of systematically interpreting and understanding the semantics of data patterns which have been recorded in security surveillance or health care applications.

Understanding and interpreting human behaviours based on heterogeneous sensor analysis, have observed competitive challenges. To address these problems, advanced image and signal processing technologies such as neural network, fuzzy logic, probabilistic estimation theory and statistical learning have been investigated.

In this paper, an approach will be presented, based on the fusion of the observations given by four different methods of classification, using a multi-modal heterogeneous sensory framework. The data fusion procedure will be accomplished by means of Concurrent Hidden Markov Models and corresponds to the high-level classification step, from a two layered classification process. This paper is structured as follows: in Section 2 we introduce the four methods in the low-level classification layer, in Section 3 the high-level classification layer is presented. In section 4 a summary of the implementation and results, for the smart home scenario, is presented. Sections 5 and 6 present the conclusions and future work, and references, respectively.

2. LOW-LEVEL CLASSIFICATION IN A GLANCE

In this section we present an introduction to each of the four methods used in the low-level classification stage. Thereby, in the following paragraphs we present, in a superficial way, the low-level classifiers based on Laban Movement Analysis, Crowd Analysis, Sound Analysis and the tracking using a 3D-Voxel Occupancy Grid. We let the reader to follow the references for each one of these different methods, since a more detailed description can be found in the appropriate literature.

One of the methods in the low-level classification stage of our approach for Human Behaviour Analysis (HBA) is inspired by the Laban Movement Analysis notation (LMA).

LMA is a well-known system to notate and understand human motions. It comprises five components (Body, Space, Effort, Shape and Relationship) to describe human motion in differ-

Supported by the European Union within the FP7 Project PROMETHEUS (EU-FP7-ICT- 2007-1-214901), www.prometheus-FP7.eu

ent approaches. Researchers have attempted to reformulate the LMA parameters for practical use in intelligent systems for human action understanding – see [1] for an example. Further description of this method can be found in [2].

The Crowd Analysis method uses crowd activity and crowd size observations to analyse group behaviours. Crowd activity provides information which can be used to detect anomalous crowd behaviour. Normal behaviour often corresponds to calm movements, whilst an anomalous event, however, is likely to be accompanied by more rapid movements. The level of crowd activity is measured by using the optical flow, which estimates the relative motion between consecutive images. The inputs to this classifier are the optical streams from the visual and thermal infra-red sensors. Although this method is typically applied to outdoor scenarios, its application in the Smart Home case gives information about the overall motion in the scene.
The crowd size can be used to get further information or warnings for fights, attempted robbery and risk of riots.
Further description of this work can be found in [3].

The Sound-based Analysis method contributes significantly to the final classification of the behaviour. The recordings, in the PROMETHEUS datasets, contain rich information as regards different types of typical and atypical sound events, while a high degree of variation exists between samples of the same category. Further description of this work can be found in [4].

The fourth method in the low-level classification stage uses a 3D Voxel Occupancy Grid tracker that allows the reconstruction of the 3-dimensional shapes of objects in multi-view scenarios. To this end, the scene is quantized to a three dimensional voxel occupancy grid. The occupancy of each of the voxels is determined by the joint observations of all available visual sensors. This method allows generating accurate 3D detections, as well as an accurate reconstruction of the scene. All the processing is done in 3D and therefore allows reasoning in a 3D environment as opposed to just flat 2D camera images. This accurate 3D reconstruction is used as the basis for event detection. To this end, we employ a new joint state tracking algorithm, which is not only capable of tracking people, but also allows detecting events simultaneously. Further description of this work can be found in [5].

3. HIGH-LEVEL HUMAN BEHAVIOUR ANALYSIS

In the second stage of Human Behaviour Classification, we proceed with a high-level classification step, using the observations received from the low-level classification stage as input.
The full system therefore consists of a hierarchical probabilistic framework which yields a high-level classification of actions, denoted as ”events”, incorporating prior knowledge on the scenarios being analysed.

3.1. Motivation for the use of Hidden Markov Models

The problem of sequences recognition with noisy data could be addressed by two different approaches that are symbolic methods and statistical pattern classification methods.
In the symbolic approach, the problem is approached as a logic inference process. A knowledge based system uses rules to build a representation of each type of sequence to recognize. These rules could be given by an expert or found by data analysis. This approach has the important ability to build models that represent human interpretations of the sensor data, but they have difficulties to deal with noisy data.
In contrast, a statistical pattern-classification system attempts to describe the sequence of noisy data as a random process. The recognition process consists of the association of a new sequence with a model of the feature to identify. This approach exhibits the important property of dealing with noisy data, but its main disadvantage is the obscure nature of the learned black box system.
Markov models have the advantage of combining the strengths of both the symbolic and the pattern-classification approaches. Hidden Markov Models (HMM) are widely employed in the field of computer vision to recognize gesture or human behaviour [6]. In these applications, the observation variables are features extracted for video data.

3.2. Concurrent HMM architecture for behaviour recognition

For behaviour recognition, we are interested in detecting the current behaviour amongst *N* known behaviours (i.e. the behaviour library). For this purpose, we propose to use a Concurrent HMM [7] architecture.
A concurrent hidden Markov model is composed of several Hidden Markov Models, each one describing one class (see Fig. 1). To summarize, the concurrent HMM centralises the on-line update of the behaviour belief and contains:

1. The set of HMMs representing basic behaviour library;

2. The transition between behaviour models that could be either defined by hand (by an expert), or learned from annotated data.

3.3. Human behaviour learning

In our application, each HMM composing the concurrent hidden Markov model describes a human behaviour and is learned using a training dataset, that is composed of labelled observation sequences that are feature extracted from the different sensors for the associated behaviour. The classical Baum-Welch algorithm is used to train the HMM matrices. The

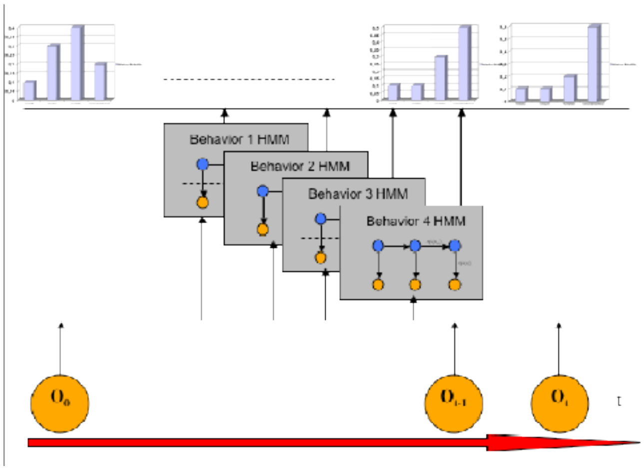

Fig. 1. Diagram for the Concurrent Hidden Markov Models approach for recognizing the different behaviours. Figure extracted from [8]

data used for training are based on knowledge about the observations. Fifty set of observations are generated based in the relations between the observations and the events, and the probabilities associated to each one.

3.4. Human behaviour recognition

As previously emphasized, the concurrent Hidden Markov model is used to recognize on-line or off-line the current behaviours amongst *N* known behaviours. This is easily performed by finding the HMM *M* that maximizes $P(M|O_{t-n}, ..., O_t)$ for the off-line case or $P(M|O_t)$ for the on-line case). It is important to note that a difficult problem is to handle the case where the behaviour to recognize is not in the training dataset (i.e. the rejection problem in the classification community). A way to solve this problem is to have a threshold on the probability of the model given the observation $argmax_M(P(M|O_{t-n}, ..., O_t))$ for the off-line case and $argmax_M(P(M|O_t))$ for the on-line case). This threshold allows to not classify observation sequences that do not correspond to the different HMMs within the concurrent HMM.

4. IMPLEMENTATION AND RESULTS

The data used in this paper has been recorded during the 7th Framework EU Project PROMETHEUS [9].
The PROMETHEUS database comprises a collection of dataset composed by a group of recordings from different sensors. These group of recordings for each scene provide a heterogeneous dataset that can provide data for all the four methods mentioned above. The PROMETHEUS database can be found at http://paloma.isr.uc.pt/DataCollectionDB/prom.
The database contains two recordings which, besides others, contain several instances of the falling down events.
The scenario was recorded using five cameras, with a total of six image channels. These include three visual channels from CCTV cameras, one thermal channel from a thermal camera, as well as a near infra-red (NIR) channel and a range channel from a Photonic Mixer Device (PMD). We divide the six channels into two different kinds of data sources (intensity and range).
The HMM models implemented for the smart home scenario were focused on detecting six different events, namely: Watch television, Fall down, Enter, Exit, Discuss, Panic. The observations used in the HMM's are provided by the LMA, Crowd Analysis, Sound Analysis, and from the 3-D Voxel Occupancy Grid tracker algorithms, expressed as a probability for the set of low-level actions (obtained as a result from the low-level classification stage). Each HMM fuses this set of evidences in order to determine the log-likelihood of the event to happen (the event with the smallest negative log-likelihood is the "winner"). The model proposed is composed by six HMMs, each one modelling its own event. The observations used in the model are binary and based in the data from the low-level classification outputs. We will have the following observations: Increasing, Decreasing, Low, High distance to entrance door from 3D-Voxel Occupancy Grid tracker O1, O2, O3, O4; Normal, Increase, Intense Activities from Crowd Analysis O5, O6, O7; Normal, Atypical Sounds from Sound Analysis O8, O9; Standing, Falling Down, Sitting from 3-D Voxel Occupancy Grid tracker O10, O11, O12; Standing, Walking, Running, Falling Down from Laban Analysis O13, O14, O15, O16.
Due to the quality of the events, the numbers of states are defined as 2 hidden states.
The observations are presented to each HMM, and the smaller negative log-likelihood associated to it is the winner.
The observations used to obtain the negative log-likelihood using the Viterbi algorithm were composed by data acquired during one second.
Fig. 2 and Fig. 3 illustrate the results obtained to one scene in the smart home scenario, where a fall occurs. Fig. 2 is the result obtained to one instant before the falling down event is acknowledge by the CHMM; and Fig. 3 is the moment that the CHMM acknowledge the falling down event.

5. CONCLUSIONS AND FUTURE WORK

With this work we have demonstrated that probabilistic techniques can be successfully used to deal with Human Behaviour Analysis. Furthermore this work focused on Human Behaviour Analysis (HBA), using a multi-modal heterogeneous sensing environment, which raised interesting challenges to be overcome. To overcome these challenges we considered an approach that performs HBA using a two step classification, using Concurrent Hidden Markov Models in the highest-level stage of the classification process to recognize between events. Therefore, the scalability of this approach is ensured due to

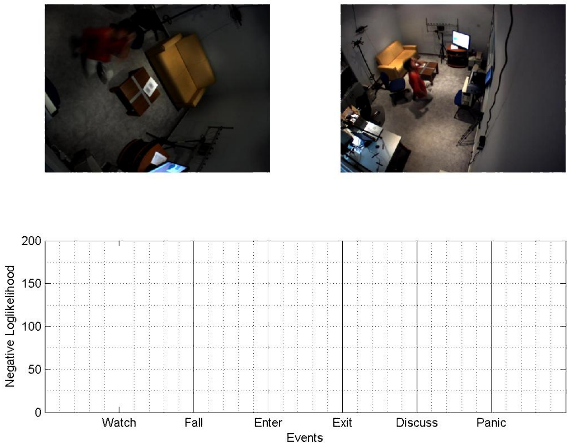

Fig. 2. Results obtained up until the moment before the falling down event. The negative log-likelihood to all events is infinite.

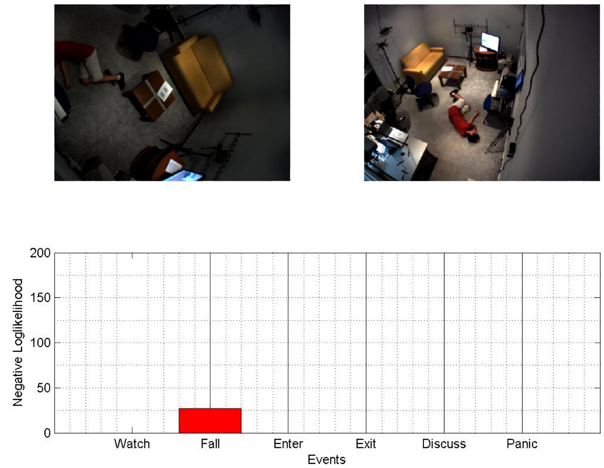

Fig. 3. Results obtained at the moment when the Falling Down event is detected by the HMM, with a small negative log-likelihood shown in the bar graph.

the CHMM characteristics. This means that for detecting a greater number of behaviours all it is required is to add the corresponding HMM to the others.
Given the high dependence between the recognized behaviours and the context where they are occurring, we will aim to integrate better contextual models with these classification algorithms, in order to enlarge the adaptability of the system.
To the moment this work was undertaken the authors couldn't find a similar approach that used fusion of multi-modal heterogeneous sensors that were working in a similar way and applied to a similar case to the one in discussion, therefore the lack for comparison results with other approaches.

6. REFERENCES

[1] Joerg Rett, Jorge Dias, and Juan-Manuel Ahuactzin, "Laban movement analysis using a bayesian model and perspective projections," in *Brain, Vision and AI, ISBN: 978-953-7619-04-6*, 2008.

[2] K. Khoshhal, H. Aliakbarpour, J. Quintas, P. Drews, and J. Dias, "Probabilistic lma-based classification of human behaviour understanding using power spectrum technique," in *13th International Conference on Information Fusion*, 2010.

[3] P. Drews, J. Quintas, J. Dias, M. Andersson, J. Nygards, and J. Rydell, "Crowd behavior analysis under cameras network fusion using probabilistic methods," in *13th International Conference on Information Fusion*, 2010.

[4] S. Ntalampiras, I. Potamitis, and N. Fakotakis, "On acoustic surveillance of hazardous situations," in *International Conference on Acoustics Speech and Signal Processing*, 2009.

[5] D. Snow, P. Viola, and R. Zabih, "Exact voxel occupancy with graph cuts," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2000, p. 345352.

[6] Wilson A.D. and Bobick A.F., "Parametric hidden markov models for gesture recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, vol. 21(9), pp. 884 – 900.

[7] A. Benveniste, E. Fabre, and S. Haar, "Markov nets: probabilistic models for distributed and concurrent systems," 2003, vol. 48, pp. 1936–1950, IEEE.

[8] Hadi Aliakbarpour, Kamrad Khoshhal, Joao Quintas, Kamel Mekhnacha, Julien Ros, Maria Andersson, Stavros Ntalampiras, and Jorge Dias, "HMM-based Abnormal Behaviour Detection Using Heterogeneous Sensor Network," in *Second Doctoral Conference on Computing and Electrical and Industrial Systems, DoCEIS11*, 2011.

[9] Stavros Ntalampiras, Dejan Arsić, Andre Störmer, Todor Ganchev, Ilyas Potamitis, and Nikos Fakotakis, "PROMETHEUS database: A multi-modal corpus for research on modeling and interpreting human behavior," in *Proc. Int. Conf. on Digital Signal Processing*, 2009.