Recognition of human actions using Layered Hidden Markov Models

Serafeim Perdikis[2], Dimitrios Tzovaras[1], Michael Gerasimos Strintzis[1,2]

[1]Informatics and Telematics Institute, Thermi-Thessaloniki, Greece
[2]Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki

Abstract

Human activity recognition has been a major goal of research in the field of human - computer interaction. This paper proposes a method which employs a hierarchical structure of Hidden Markov Models (Layered HMMs) in an attempt to exploit inherent characteristics of human action for more efficient recognition. The case study concerns actions of the arms of a seated subject and depends on the assumption of a static office environment. The first layer of HMMs detects short, primitive motions with direct targets, while every upper layer processes the previous layer inference to recognize abstract actions of longer time granularities. The results demonstrate the efficiency, the tolerance on noise interpolation and the high degree of person - invariance of the method.

Categories and Subject Descriptors (according to ACM CCS): I.2.6 [Artificial Intelligence]: Learning, H.1.2 [Models and Principles]: User/Machine Systems

1. Introduction

Automatic Human Activity Recognition (HAR) has received great attention by researchers involved in human - computer interaction, due to the continuous need for smarter and more user - friendly interfaces. HAR implementations presented so far vary widely in terms of the medium of surveillance (e.g. camera, motion tracker), the target of recognition (e.g. indoor or outdoor activity), the human model and the mathematical model.

As far as the mathematical model is concerned, activity recognition methods can generally be classified into those who employ a state - space model (Bayesian Networks, Finite State Machines, Hidden Markov Models) and those who rely on pattern recognition techniques (Support Vector Machines, Neural Networks, Dynamic Time Warping, Bayes and K - means classifiers).

State - space models and especially Hidden Markov Models (HMMs) have been preferred in most cases for solving the activity recognition problem, due to their efficiency in capturing spatio - temporal dynamics of signals [LDG*04]. In this paper a layered HMM structure (LHMMs) is applied to replace the typical single - layer HMM classifier, thus facilitating the learning and inference procedures. By decomposing the inherent structure of human activity, the method manages to reduce the training requirements of the HMMs, thus enhancing the efficiency and robustness of the recognition system.

The paper is organized as follows: in Section 2 the basic ideas behind the proposed method are explained. In Section 3 implementation issues are thoroughly discussed and in Section 4 some results are presented and commented on.

2. Method Description

The key feature of the HMM recognition framework is the property that given a HMM λ, a probability $P(O|\lambda)$ can be assigned to the generation of any observation sequence O. Observation sequences can be denoted $O = O_1 O_2 \ldots O_t \ldots$, where $O_t = \{Feature_1, Feature_2, \ldots\}$ the feature vector at time slot t.

The classical single - layer approach for HMM activity recognition suffers certain limitations. Modeling actions of relatively long duration leads to long observation sequences that burden the training process and reduce the efficiency of the recognition. Besides that, extraction of a large number of activity features (e.g. multi - sensorial environments) augments the training data encumbering the inference process. These drawbacks can be overcome by implementing a

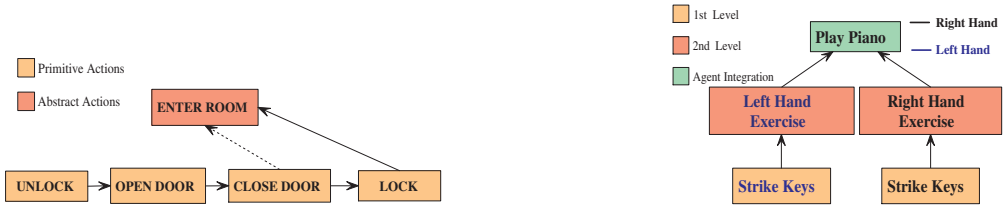


Figure 1: *Structure of human activity*

layered structure of HMMs. Layered HMMs (LHMMs) can improve the training process in two fashions. First, the high - complexity processing of low - level data is restricted to the first layer only, permitting the upper layers to process simple discrete input symbols, based on the previous layer inference. Additionally, LHMMs can achieve efficient segmentation of the parameter space, by integrating the inferential results of multiple HMMs in the same layer. Related work can be found in [OHG02], [NPVB05] and [ZN05].

The contribution of this paper lies mainly on the demonstration of the applicability of LHMMs for the Activity Recognition problem, when a person's actions (e.g. putting a stamp), rather than his "state" or "situation" (e.g. phone conversation, [OHG02]) has to be detected. Recognition of abstract actions proves to be a challenging problem, since the order of the series of events is of great importance. In order to achieve this goal, a decomposition of the structure of human action is necessary. Eventually, the application of LHMMs becomes feasible thanks to the innovative idea of exploiting two typical characteristics of the human activity:

Hierarchical and chronological structure of activity
Human actions can be classified into hierarchical levels of abstraction. The lowest level of human activity hierarchy is occupied by simple, short motions with single, direct targets, referred to as Primitive Motions (PMs). Every upper layer contains more Abstract Motions (AMs), that take place in longer time intervals, accomplish more complicated goals and reveal complex intentions. Actions at some level are composed by a sequence of actions of the previous level. In this manner, actions in successive levels are connected to each other, because every action can be described as the result of the execution of simpler actions at the previous level over some period of time. An example of this structure is shown in Figure 1(a).

Distribution of activity to multiple cooperative agents
Another inherent characteristic of human activity is the execution of composing actions by different motion agents. When a single human is considered, the role of motion agents is played by the human limbs. For instance, walking consists of periodical movements of the two legs. When a whole team is taken into consideration, then every member can be seen as an agent, whose action contributes to the fulfillment of the team's objective. The knowledge about the activity of every single cooperative agent is crucial for a reliable inference about the type of the overall activity. Figure 1(b) presents an example of how the combination of agent inferential results differentiates the final inference.

The above observations inspire a layered structure of HMM model for activity recognition. More specifically, the LHMMs recognition method is based on the following ideas: A set of N motion agents $A = \{A_1, A_2, \ldots, A_N\}$ is defined for the activity in question. A set of M_i Primitive Motions is defined for every agent A_i $(1^{st} level)$: $PM^{A_i} = \{PM_1^{A_i}, PM_2^{A_i}, \ldots, PM_{M_i}^{A_i}\}$. A set of R_i Abstract Motions is defined for every agent A_i $(2^{nd} level)$: $AM^{A_i} = \{AM_1^{A_i}, AM_2^{A_i}, \ldots, AM_{R_i}^{A_i}\}$. More layers can be added as the level of abstraction of the described actions increases.

For every layer L of an agent A_i, a bank of HMMs is assigned performing a mapping of the layer's observation sequences O_L to the actions X^L contained in this layer: $f_L : O^L \rightarrow X^L$. For the first layer, the observation sequences O^1 are sequences of feature vectors extracted by the raw input data, while the actions X^1 belong to the set PM^{A_i}. The mapping procedure f_L at every layer L implements the classical HMM recognition framework. For the second and every upper layer, the observation sequences consist of the inferential results of the previous layer over some period of time. Thus, successive outputs of some layer form the (discrete) input vectors of the next one.

At some level an integration procedure takes place, so that the overall activity can be inferred by the partial inferential results of every single agent alone. The agent integration process concerns the detection of meaningful, simultaneous, and cooperative actions among the defined activity agents. Figure 2 depicts graphically the proposed method. The advantages emerging by the application of the method include: a) the restriction of continuous observation sequences, that require laborious processing, to short sequences at the 1st layer only, through the introduction of levels of abstractions, and b) the segmentation of long feature vectors to multiple shorter ones thanks to the introduction of multiple motion agents.

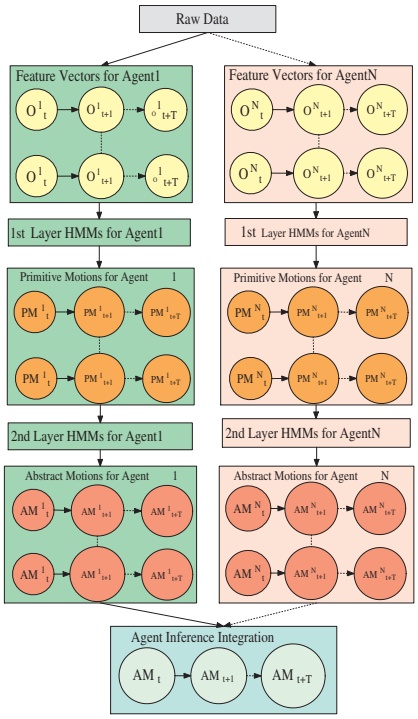


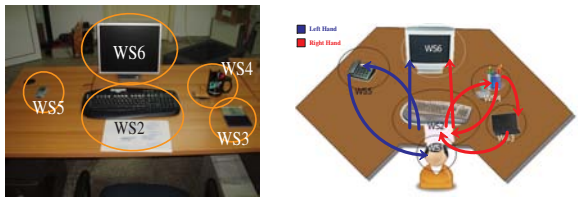
Figure 2: *Layered HMM method block diagram*

3. Activity Recognition in office environment

The functionality of the proposed method has been tested under a simple implementation scheme containing two layers. The target of recognition concerns actions of the arms in an office environment, suitable for an office awareness application: Pick Up Phone, Adjust Screen, Switch Screen On/Off, Take Pen and Put Stamp.

The implementation of the method relies on the assumption of a static office environment, where the positions of all objects on the desk and the subject's seat are relatively fixed. With regard to the analysis in Section 2, two cooperative agents are defined, namely the two arms of the subject, denoted LA and RA for the left and right arm respectively. The static office environment is divided into 6 workspaces $WS_i, i = 1,2,\ldots,6$ as shown in Figure 3(a). Workspaces can be viewed as the surrounding space of one or more objects.

The reason for introducing the static environment and the workspace definition is, that this scheme enables the bounding of the PM set for both agents to transitions between two workspaces. Formally, PMs can be denoted $LAWS_iWS_j$ or $RAWS_iWS_j, i \neq j$ respectively. Finally, 8 PMs have been defined for the first level of abstraction, 3 for the left and 5 for the right arm. In Figure 3(b), PM transitions are represented as arrows in the static environment. According to the method description, every AM of the second level is formed



(a) (b)

Figure 3: *Workspaces in static office environment and PM definition*

by a sequence of PMs of the previous level, following the natural structure of human activity. With respect to that, the final form of the implementation scheme is presented in Figure 4. It is important to underline that the distinction of the actions Adjust Screen and Switch Screen On/Off can only be achieved after the agent integration procedure dictated by the method's formulation. The subject's arms are modeled with

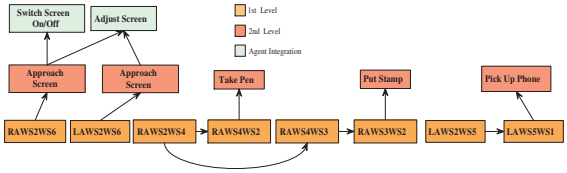
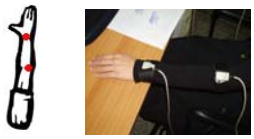


Figure 4: *Implementation scheme diagram*

two distinctive body spots, the wrist and the elbow (Figure 5(a)). The trajectories of these spots are captured by a wearable magnetic motion tracker (Ascension MotionStar©, Figure 5(b)). Consequently, the raw data produced at every single time - slot t for both agents contains 3D positions of the associated body spots: $R_t = \{x_W^t, y_W^t, z_W^t, x_E^t, y_E^t, z_E^t\}$, where W stands for wrist and E for elbow. The motion features extracted are the 3D position and the vectorial velocity, so a feature vector at time slot t can be denoted: $O_t = \{x_W^t, y_W^t, z_W^t, x_E^t, y_E^t, z_E^t, Vx_W^t, Vy_W^t, Vz_W^t, Vx_E^t, Vy_E^t, Vz_E^t\}$. The banks of HMMs were trained with motion samples



(a) (b)

Figure 5: *Arm model and wearable motion tracker*

taken by 7 different subjects using the Baum - Welch parameter estimation algorithm [Rab89]. Instead of a single HMM per PM, 5 HMMs per PM were trained, in order to capture variations on the execution time of the actions. In

the inference phase, the first layer of every agent emits every 150 msec a discrete symbol associated to the PM taking place at that time. Although the inference procedure is based on the classical HMM recognition framework, an auxiliary decision system has been employed to improve the recognition performance. Furthermore, raw data input undergoes an Euclidean distance segmentation process before being fed to the first layer HMMs, so that only segments of the testing sequence where motion has been detected are taken into consideration. The segmentation procedure increases the speed of inference and eliminates false alarm errors.

The concatenation of the first layer inference symbols over longer periods of time, form the observation sequences of the second layer HMMs. It should be noted that in case of an absolutely accurate inference at the first layer, simple Finite State Machines instead of HMMs could be used at the second layer to detect the desirable sequence of PMs that form an AM. In fact, the first layer's inference proved to be prone to wrong decisions over short time intervals. For this reason, the symbols emitted by the first layer are treated as observation symbols of the "hidden" state, which represents the actual PM currently executed. Accordingly, second layer HMMs have been trained in a novel heuristic manner (direct specification of the HMM parameters) to recognize trivial, frequent mistakes of the first layer, thus "correcting" the inferential results and enhancing the robustness of the system.

4. Results and Conclusions

Table 1 presents the results acquired by testing the classifier with 5 subjects instructed to naturally perform 5 repetitions of the target actions with short time gaps in between. Subjects of both sexes and varying physical features were picked for both training and testing. Recognition rates were noted over the 25 samples of each action. The results demonstrate a recognition rate over 80% for all the actions in question.

In the previous case, measurements are assumed to be noiseless due to the accuracy of the magnetic tracker. However, noisy measurements are expected in applications where motion capture is performed using computer vision techniques. In order to determine the influence of this limitation, the effects of Gaussian noise interpolation on the raw data were investigated, revealing the immunity of the LHMM system to noise with SNR > 10 (Figure 6). Concluding, the presented implementation demonstrates that LHMMs can be successfully employed for the recognition of human actions, achieving more efficient training, reliable inference and improvement of the system's robustness. Additionally, training and testing the classifier with different individuals reveals a certain person - invariance of the classifier. Our current work addresses the static environment limitation through unsupervised learning. Possible displacements of the workspaces can be "learned" by retraining the first layer in an automatic, unsupervised fashion.

Table 1: *Recognition rates*

Abstract Motion	Recognition rate
Pick Up Phone	100%
Adjust Screen	100%
Switch Screen On/Off	100%
Take Pen	80%
Put Stamp	92%

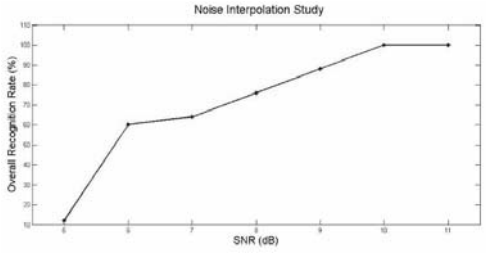


Figure 6: *Performance of the proposed method in the presence of measurements noise*

References

[LDG*04] LEO M., D'ORAZIO T., GNONI I., SPAGNOLO P., DISTANTE A.: Complex Human Activity Recognition for Monitoring Wide Outdoor Environments. In *ICPR '04: Proceedings of the 17th International Conference on Pattern Recognition, (ICPR'04) Volume 4* (2004), IEEE Computer Society, pp. 913–916.

[NPVB05] NGUYEN N. T., PHUNG D. Q., VENKATESH S., BUI H.: Learning and Detecting Activities from Movement Trajectories Using the Hierarchical Hidden Markov Models. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2* (2005), IEEE Computer Society, pp. 955–960.

[OHG02] OLIVER N., HORVITZ E., GARG A.: Layered Representations for Human Activity Recognition. In *ICMI '02: Proceedings of the 4th IEEE International Conference on Multimodal Interfaces* (2002), IEEE Computer Society, p. 3.

[Rab89] RABINER L. R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE 77*, 2 (1989), 257–286.

[ZN05] ZHANG X., NAGHDY F.: Human Motion Recognition through Fuzzy Hidden Markov Model. In *CIMCA '05: Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce Vol-2 (CIMCA-IAWTIC'06)* (2005), IEEE Computer Society, pp. 450–456.