Recognition of Continuous Activities

Abstract. The recognition of continuous human activities performed with several limbs is still an open problem. We propose a novel approach for recognition of continuous activities, which considers the direction change between frames to track the motion of several limbs, uses a Bayesian network to recognize different activities. The approach presented can recognize activities performed at different velocities by different people. We tested the model with real image sequences for 3 different activities performed on a continuous way.

Keywords: Bayesian networks, human activity recognition, multiple motions, continuous activity recognition

1 Introduction

Performing an activity can imply that a human moves one or several limbs at the same time. In the case of sports, for example, when an athlete throws a ball he or she has to move his arms and legs at the same time. A good throw requires coordination between the limb movements. In order to recognize activities like this, we need to use a model able to work with information from different sources. Given a sequence of images we can extract information about the activity and then recognize it. Therefore, we require good models to represent and classify the activities. In this work we propose a human activity recognition model based on Bayesian networks. A Bayesian network realizes the classification of activities by associating the limb movements with different activities. The proposed model includes several outstanding aspects:

- Characterization of the activity based on its global trajectory in order to overcome occlusion or noise problems.
- Recognition of activities performed moving several limbs at the same time.
- Recognition of activities performed with different velocities.
- Representation and recognition of different activities using just one model.

- Recognition of activities performed on a continuous way.

The relevant previous work is commented in the next section. We describe our model to represent activities in section 3. The recognition network is presented in section 4. Experimental results are summarized in section 5. In section 6 we give some conclusions and directions for future work.

2 Related Work

When we want to recognize human activities we have to consider two aspects: representation and recognition. The former has been dealt in different ways. The human body and the activities have been modeled like silhouettes [1] and like temporal templates [2]. However, by trying to represent an activity with a silhouette pose, the silhouette extraction is a problem when shadows exist or there are many objects in the environment. In the representation with temporal templates, it is considered the global movement of the body, so that a change in the movement of some limb, no previously defined, will indicate a different activity.

The recognition process has been done using two main approaches: non-probabilistic and probabilistic. The work of Ayers and Mubarak [3] is an example of the non-probabilistic approach. They recognize human actions in an office environment. Their system works if it has a previous description of the environment. It can not work in unpredictable environments. In order to overcome the uncertainty or the changes in the environment the probabilistic approach has been developed. The most used probabilistic models in human activity recognition are hidden Markov models [4], [5]. Alternative probabilistic models are Bayesian networks [6]. Their application in computer vision has focused in the description of high level information. Probabilistic methods can deal with uncertainty and changes in the way the activity is performed.

In activity recognition, some situations make the recognition process more difficult. Sometimes one limb hides the other (auto-occlusion). This makes more difficult to follow the activity and the recognition process fails because of missing information. Another situation that we have to consider is that people perform the activities with different velocities and in different ways. All the previous works that recognize activities with hidden Markov models define one model for each activity. Previous works do not analyze the coordination between the movements of different limbs (multiple motions). So far the recognition of activities performed in a continuous way has not been achieved. In the next sections we describe a model that deals with the previous problems.

3 Activity Modeling

The first step in human activity recognition is to represent the activities that we want to recognize. Basically, we consider that when a person performs an activity we can have a global trajectory of each limb describing the movement. The global trajectory of a limb is its position sequence (X,Y,Z) obtained when the activity is performed just once. The beginning and the end position of the activity can change. We initially consider activities performed with the arms only. We use a color landmark in the wrists to know what limb is moving and apply a color detection process to get each landmark (figure 1). Then we extract its center of mass (we initially consider its position in just one plane, X-Y) to get the wrist position in each frame and track the activity trajectory. In this way the position of each limb is know every instant.

Working only with the X, Y positions of the landmarks' centroids is very susceptible to the distance between the person and the camera. Therefore, we calculate the direction between two consecutive frames using the X, Y positions. The directions are discretized in 8 sections as shown in the figure 2. When there is no movement between frames we assign a value of zero. The problem here is how to determine how many frames are required to represent an activity since an activity is performed with different velocities by different persons or even by the same person. The activities that we are considering can be performed in at most one second. The frame rate employed is 15 frames per second. The movement of each limb has a direction sequence describing an activity. The most important aspect in sequence is the direction changes. Based on statistics from different activities by different people, we obtained that in a sequence of 15 frames at most there are 7 direction changes. So in a window of 15 frames we look for at most 7 direction changes (figure 3). These directions are used in the recognition network, presented in the next section.

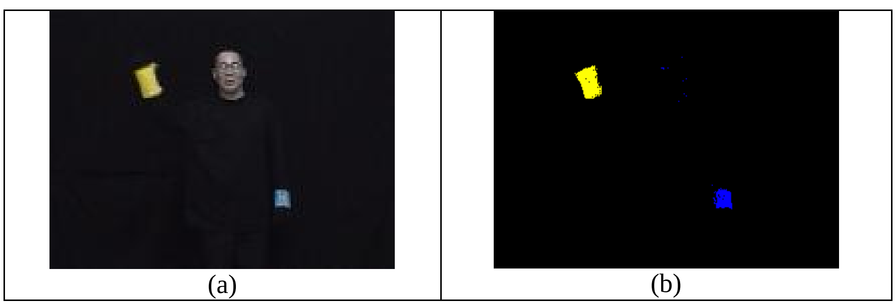

Fig. 1. (a) Original image with landmarks in the wrist. (b) Landmarks after using a detection color process

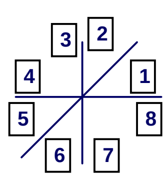

Fig. 2. Discretized directions

5	5	5	4	4	4	8	8	8	1	1	1	2	5	5
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Fig. 3. Example of a window with direction changes. The first row shows direction changes. In this example there are 6 direction changes (5, 4, 8, 1, 2, 5). The second row has the number of frames

4 Recognition Network

Activity recognition starts by capturing an image sequence in which one person is performing some activity. The model in figure 4 shows a Bayesian network [7] in which the root nodes represent the activities that are going to be recognized and the leaf nodes represent the direction changes (observations) gotten from a window of 15 frames. If in that window there are less than 7 direction changes we instantiate with a zero the nodes without direction change. Considering the direction changes we are representing activities performed with different velocities because the activity is characterized by its direction changes no by its velocity. We are considering that the activities to be recognized can be performed with one or both arms. The node values at the lower levels change according to change of the activity direction reflecting the evolution of the limb motion. This network can be used to complete the trajectory in the case there is missing information (occlusion) or noise, providing the most probable trajectory. The network is able to track the motion of each limb at the same time doing a multiple motion analysis. The nodes in the middle give information about the type of movement of each arm.

The leaf nodes in the network are instantiated by the observations in the correct window, and then the probabilities are propagated [8] in the network in order to get the posterior probabilities of the root nodes. The root nodes have two possible values (*Yes*,*No*) for each activity. In this way we can recognize several activities performed at the same time. After propagation, several nodes can have a high probability in its value "*Yes*". Also, the network is able to recognize when the movement of several limbs form an activity. This implies that there is coordination between the limb motions.

The Bayesian network model is applied in a window of 15 frames, and then displaced 3 frames to realize a continuous recognition. The activities to be recognized can take less than 15 frames. We do not need to mark the beginning and the end of the activity. The frames that generate a high probability in one or more of the root nodes indicate when an activity considered in the model is performed.

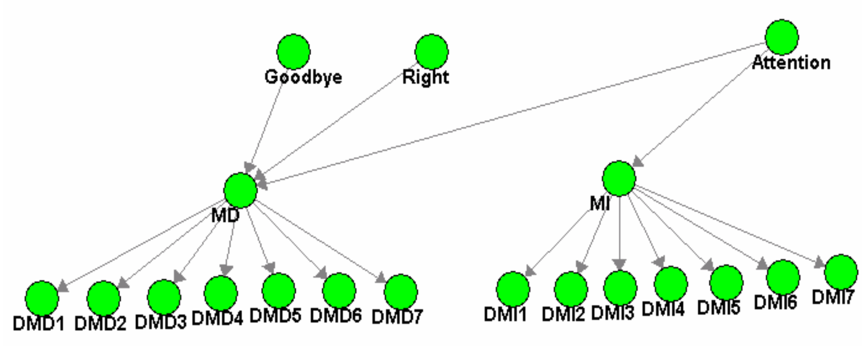

Fig. 4. Recognition network. The root nodes represent the activities to be recognized. The middle nodes contain information about the movement of each arm. The leaf nodes have the direction changes

5 Experimental Results

We trained the network with videos of 5 persons performing 3 types of activities: *goodbye*, *move right* and *attract the attention*. *Attract the attention* is a compound activity in which the movement of both arms is required. This involves the recognition of multiple motions. Every person performed the activities with different velocities so the number of direction changes needed to represent the global activity varied. We tested the activity recognition system using continuous activities. For example, figure 5 shows part of a sequence that had 378 frames where the person performed the three activities in a continuous way. The system was able to recognize correctly the activities *goodbye* and *attract the attention* in different sections of the test sequences. In some sections the system did not recognized the activity, but it did not give a wrong classification, neither.

6 Conclusions

We have developed a model that can recognize compound activities (performed with several limbs at the same time). It can recognize different activities performed continuously at different speeds and by different persons by using a window to get the direction changes of global activity trajectory. We made an analysis on how many direction changes are enough to have a general activity representation, and concluded that a reduced number of changes are sufficient for representing an activity and for its recognition. We developed a system that uses a Bayesian network for continuous recognition. With only one model we were able to model and to recognize different activities.

Future work will test the recognition network and the window displacement in order to improve the continuous recognition.

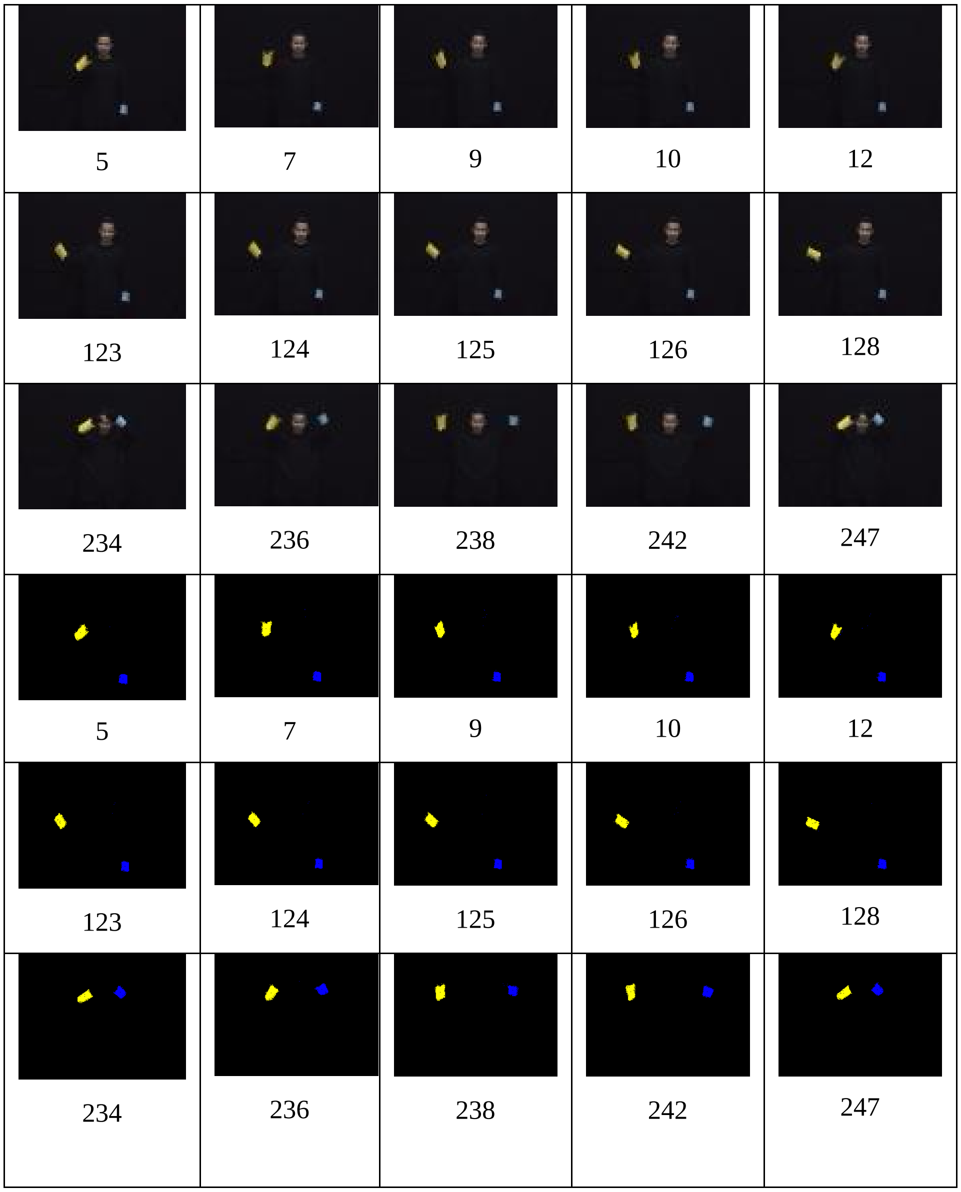

Fig. 5. In the 3 first rows, an image sequence is showed with a color landmark in both wrists. In the 3 bottom rows, the color landmark is depicted after applying a color detection process on the original images. The number below is the frame number in one continuous sequence of the activities “goodbye (5-12), move right (123-128) and attract the attention (234-247)”

References

1. Ismail Haritaoglu, David Harwood, and Larry S. Davis, “Ghost: A Human Body Part Labeling System Using Silhouettes”, Fourteenth International Conference on Pattern Recognition, Brisbane, August, 1998.
2. James W. Davis and Aarón F. Bobick, “The Representation and Recognition of Action Using Temporal Templates,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR’97), 1997.
3. Douglas Ayers and Shah Mubarak, “Monitoring Human Behavior in an Office Environtment,” Interpretation of Visual Motion Workshop, CVPR’98, june 1998.
4. Christoph Bregler, “Learning and Recognizing Human Dynamics in Video Sequences,” Proceedings IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, june, 1997.
5. Junji Yamato, Jun Ohya and Kenichiro Ishii, “Recognizing Human Action in Time-Sequential Images Using Hidden Markov Model,” IEEE, 1992.
6. Stephen Intille and Aaron Bobick, “A Framework for Recognizing Multi-Agent Action from Visual Evidence,” M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 489, 1999.
7. Richard Neapolitan, “Probabilistic Reasoning in Intelligent Systems: Theory and Algorithms,” New York: Wiley, 1989.
8. Steve B. Cousins, William Chen, Mark E. Frisse, “CABeN: A Collection of Algorithms for Belief Networks” Medical Informatics Laboratory, Washington University, 1990.