Multi-Level Modeling and Recognition of Human Actions Involving Full Body Motion

Luc Emering, Ronan Boulic, Selim Balcisoy, Daniel Thalmann

LIG - Computer Graphics Lab, Swiss Federal Institute of Technology, Lausanne, CH-1015 Switzerland

{emering,boulic,ssbalcis,thalmann}@lig.di.epfl.ch

Abstract

A multi-level model of human actions involving full body motion is proposed. Actions are considered as the combination of action primitives at different levels : motions of the center of mass, end effectors, and final postures of the virtual skeleton. The model is especially designed for fast action recognition from the input of the human motion at the joint level.

1. Introduction

We consider here the modeling and recognition of "every day's life" human actions as opposed to the command coding through human motion or postures (e.g. sign langage, sport referee, or alternate communication interface [RP+94]). Our input is the full body motion thus providing a very general framework to our approach. Previous work relied on image analysis to derive interaction between a real person and a synthetic dog [MDBP95]. However only visible body parts can be retrieved in real time which prevents a significant repertoire of actions to be identified. We exploit a magnetic motion capture system yielding the user's joint values in real-time [MBT96]. In this paper, we propose a multi-level action model dedicated to the fast recognition of a large range of actions. Its robustness is evaluated in [EBBT97] for real-time behavioral interactions with virtual humans.

2. Action Modeling

The main purpose of the action model is to identify the on-going action at the lowest cost in computing time and memory allocation. For actions characterized by a clear final posture, e.g. sitting, we want to predict that action before the final posture is stable. For actions characterized by a cyclic motion pattern, e.g. walking, we want to identify it at a high level over only a fraction of the cycle. Furthermore, we want to be free of any assumption for the beginning of an action, e.g. one may sit by beginning from a standing posture or from a lying posture.

The proposed action model characterizes an action in term of the body variations over a short period of time, i.e. the *gesture* level, or in terms of the expected final posture, i.e. the *posture* level, or both (Fig. 1). For efficiency reasons the gesture level encodes only the motion of two fundamental body features : the center of mass and the end effectors. The posture level integrates also the joint data in the action definition (Fig. 1). Each of the five levels of this characterization may be defined by a so-called *action primitive* or a boolean expression of action primitives that we now decribe in further details.

Autonomous Agents 97, Marina Del Rey, California USA
© 1997 ACM 0-89791-877-0/97/02 ..$3.50

Action : A full body action can be characterized at up to five levels	Gesture	1	Center of Mass
		2	End Effectors
	Posture	3	Center of Mass
		4	End Effectors
		5	Joint values

Figure 1. : The five levels characterizing a full body action

Three coordinate systems are necessary to describe the positions and motions of the CoM and the EEs. The first frame is the *global coordinate system* (in short GCS). The second is the *body coordinates system* (in short BCS) located at the spine root. Its main axis are relative to the body, i.e. Lateral, Frontal and Up (Fig. 2). The *floor coordinate system* (in short FCS) is located at the vertical projection of the spine root on the floor level. It reuses the GCS Vertical axis, normalizes the BCS lateral axis after projection on the floor and constructs the third axis with the following cross product: FCS_Frontal = FCS_Up x FCS_Lateral.

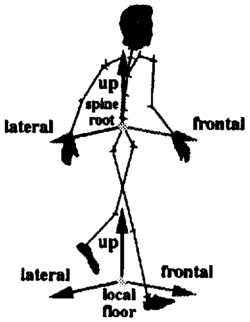

Figure 2 : Reference frames

The position and velocity of the CoM are expressed in the FCS. We use an approximated body mass distribution for that purpose. The EEs positions and velocities are expressed in the BCS. The head is a special end effector as we consider its orientation rather than its position (we examine the 'look-at' vector attached to the head).

Any action must be defined by at least one action primitive in one of the five levels. For the gesture level, we consider the motions of the CoM and EEs only along the main axis of the BCS and FCS. The velocity direction descriptors are : *up*, *down*, *forward*, *backward*, *left*, *right* and *not_moving*. An action primitive at that level has the form `(CoM, velocity direction)` or `(EE, velocity direction)`. Boolean operators can be used to define complex motion pattern as in walking :

'walking' = ((spine_root, forward) AND (left foot, forward))
OR ((spine_root, forward) AND (right foot, forward))

At the posture level, an action primitive specifies the final posture of the action. The posture is expressed in the joint space but also in the EEs position space (BCS) and the CoM position space (FCS) in order to have additional filtering levels at the recognition stage (Fig. 1 & 3.). Due to anatomical variations from one performer to another, all the spatial data are normalized by the total height of the body.

3. Action Recognition Algorithm (ARA)

The recognition process exploits the multi-level action model in order to perform in real-time (at least 10 frames/sec).

The principle is simple (Fig. 3). First, the Candidate Action Set (in short CAS) is initialized with the whole action database. Then, the motion capture system provides the user's posture directly in terms of joint angles [MBT96]. This information drives the action selection process at the five levels of the action model. The ARA retains only those actions which match the current action's characteristic or which do not define any action primitive at that level. The interest is to have low dimension matching at the higher level, with a large action database, while matching between high dimension data is made only on a small number of candidates (Fig. 3). Futhermore, the hierarchical approach overcomes the limitation resulting from the extreme simplicity of the matching algorithm while allowing real-time performance.

Multi-Level Gesture Recognition

For each new body posture sample the body movement is analyzed to derive the average velocity of the CoM and EEs over a short period of time. The recognition algorithm proceeds in a uniform fashion for all the levels of the action model. For each level, first the actions that do not define any action primitive for that level just bypass it. Second, it evaluates the action primitives triggered by the current motion. Then the remaining candidate actions exhibiting these action primitives in their definition (CoM level) or fulfilling their boolean expression (EEs level) are selected. If no action remains from the candidate set, the algorithm stops and reports an "unknown action" as output. Otherwise, the resulting CAS becomes the input of the next lower level. When no motion is detected, the gesture level reports it.

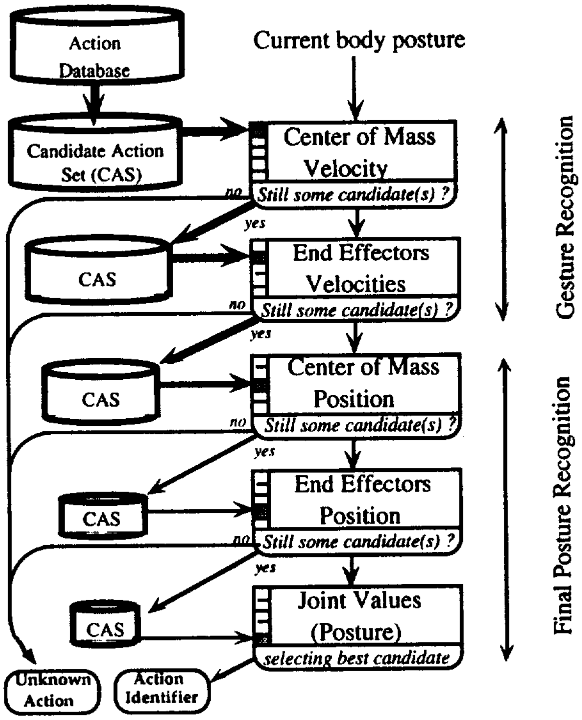

Figure 3: Multi-level action recognition process with the five successive matching units

Multi-Level Posture Recognition

The algorithm also takes advantage of the three levels of the posture model : CoM, EEs and skeleton joints. The main difference with the gesture recognition comes from the selection process. As it has the same structure for the levels 3,4 and 5 (Fig. 1) we illustrate it only for the CoM. For the actions remaining in the CAS, the algorithm computes the (squared) distances between their final posture CoM and the measured posture CoM. Then it keeps the actions for which the distance is smaller than the following radius :

$$R = (\mathrm{Min} + (1-S)*(\mathrm{Max}-\mathrm{Min}))$$

Where Min is the smallest distance found, Max is the largest distance found, and S is a selectivity parameter within [0,1]. The same algorithm is applied to the EEs level with possibly a different selectivity factor. Finally, at the joint value level it retains only the posture with minimal distance and considers it as the ultimate recognition result.

4. Discussion

The present methodology is very general and can be applied to identify the actions of other living beings on-line or off-line. Any motion capture system can be used as long as it provides the major body mobilities. Presently, magnetic sensors systems perform better in real-time for full body posture tracking. The multi-level nature of the action model allows to exploit an extremely simple matching algorithm while allowing real-time performance (see [EBBT97]).

The output of our identification can be used as an intuitive interface with synthetic agents in virtual environments ; this is explored in [EBBT97]. Due to a clear distinction between the different levels of action primitives, the action database can be easely extended to meet specific needs. Presently it is limited to the recognition of one action at a time ; we are investigating the identification of simultaneous actions, usually a principal and a secondary in terms of body motion (e.g. walking + making a phone call).

5. Acknowledgments

We are grateful to the reviewers for orienting our focus on the original part of this work. The research was partly supported by the Swiss National Foundation for Scientific Research.

6. References

[EBBT97] Luc Emering, Ronan Boulic, Selim Balcisoy, Daniel Thalmann "Real-Time Interactions with Virtual Agents Driven by Human Action Identification", First ACM Conf. on Autonomous Agents'97, Marina Del Rey, 1997

[MBT96] Molet T., Boulic R., Thalmann D. (1996), *A Real Time Anatomical Converter For Human Motion Capture*, Eurographics workshop on Computer Animation and Simulation, R. Boulic & G. Hegron (Eds.), pp 79-94, ISBN 3-211-828-850, Springer-Verlag Wien

[MDBP95] P. Maes, T. Darrell, B. Blumberg, A. Pentland, *The ALIVE System: Full-body Interaction with Autonomous Agents*, Computer Animation '95, Geneva, Switzerland

[RP+94] Roy D. M., Panayi M., Foulds R., Erenshteyn R., Harwin W.S., Fawcus R., *The Enhancement of Interaction for People with Severe Speech and Physical Impairment through the Computer Recognition of Gesture and Manipulation*, Presence, Vol.3 No. 3, Summer 1994, 227-235, MIT