eNTERFACE'12
Full project proposal

Human motion recognition based on videos

Leader : De Beul Dominique - PhD student

Candidate Participants :

1 Abstract

Imagine, a large video surveillance database or a large video dance clip database. For statistical reasons, you want to know for a particular crossroads how many people are going in a specific direction for a period of time or you want to know, how many people falls on escalators in a shopping center per day. As choreographer, you are looking for a typical footstep movements (body movements) or a dance trajectory plan. How can we help you ?

In this project we would like to develop a system that will search in videos surveillance and videos dance clip some movements and trajectories with the help of the human motion recognition.

2 Project Objectives

The main objective of this project is to design and implement a system that will search in videos surveillance and videos dance clip some movements and trajectories by using human motion recognition.

In our scenario, the user, via a web interface will be able to ask to our system the trajectories or the movements by using pre-defined lines, curves, etc. Later, we can propose to the user to draw by himself the trajectories or movements that he wants to find in our video dataset. Most of the work will be done offline such as segmentation of human bodies and key poses detection for tracking the trajectories and detecting movements. A data structure will be built to optimize the search and some a priori models will be implemented to guide the tracking module. Ontology [10] will help to search out which body movements are done in the videos. This will be formalized using the Ontology Web Language (OWL) and the ontology editor and knowledge-base framework : Protégé [1].

The objectives of the project are the design and the implementation of the following modules:

- HMI - user interface
- Tracking - segmentation, key poses detection
- Database structure
- A priori model
- Ontology

[1] http://protege.stanford.edu

3 Background information

Human motion recognition (HMR) also called human action (or activity) recognition [3] has applications in many domains such as visual surveillance, content-based video database query and retrieval, human-computer interaction [8]. HMR identifies the actions performed by body movement of human beings and is also the latest stage of the human motion capture process. This process was proposed by Moeslund [5, 6], it consist of a general structure for systems analysing human body motion (fig.1). A motion capture definition is given by Menache [4] : "Motion Capture is the process of recording a live motion event and translating it into usable mathematical terms by tracking a number of key points in space over time and combining them to obtain a single three-dimensional representation of the performance."

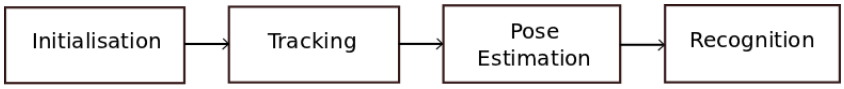

Figure 1: A general structure for systems analysing human body motion [5].

The first stage of this taxonomy (fig.1) is the initialisation, which allows to take pre-defined actions to be able to interpret the scenes which will be analyzed in the next stages. These actions can be camera calibration, adjusting thresholds, capturing reference images for the tracking stage, etc. The tracking stage will identify correctly a moving target (human) coming from a camera or a video stream. In motion capture, segmentation or extraction of information is commonly done by background subtraction that discriminate the background (scene) from the foreground (target). A review of some background subtraction techniques are proposed by Piccardi [7]. One of the difficulties of segmentation is the shadows elimination from the human silhouettes. This can be avoided by using invasive techniques such as markers or special clothing. The disadvantage is that markers can restrict natural movements and require time consuming manual post-processing of captured sequences, which making less suitable for real time applications [1]. Another problem with the segmentation algorithms is the processing time for each frames [2] which is too high. This can be an issue, especialy for real time applications. One solution is to use Graphics Processing Unit (GPU) systems. The Pose Estimation describes the posture of a human subject with help of a model or an abstraction of the human form. The *a priori* model guides the tracking by using some a priori information of a human. When no models are available, prediction or estimation algorithms are used to find feature such as position, velocity, shape, texture and colour. Recognition stage will labeling image sequences with action labels [8]. Most of the time, it is a classification problem. Several approaches are proposed in the literature [9]: Support Vector Machine

(SVM), Hidden Markov models (HMMs), Dynamic time warping, Bayes classifier, etc.

4 Detailed technical description

4.1 Technical description

We will focus on the following topics:

- **HMI module** : designing and implementing the interface between users and the system. We would like to find a good manner to represent movements and trajectories.
- **Tracking module** : selecting and implementing a background subtraction technique which is suitable for both datasets.
- **Database module** : elaborating a structured storage. We would like to compare the classic relational database to the NoSQL database.
- **A priori model module** : developing and implementing an a priori model for both type of videos.
- **Ontology module** : defining the ontology purpose, conceptualization, formalization and validation are the steps that we will follow.

4.2 Resources needed

As we believe the workshop duration will be too short to complete the project, some modules will be prepared in advance. The resources needed for the project are the following :

- Video surveillance and video dance clip databases should be ready before the project. Additional data can be collected foe adaptation and test purposes.
- A dedicated computer for the demo application is required.

We would also require usual team work facilities, such as a projector, a whiteboard, a meeting space.

5 Work plan and implementation schedule

	Week 1	Week 2	Week 3	Week 4
HMI module		x	x	
Tracking module	x	x		
Data module	x	x		
A priori module		x	x	
Ontology module			x	x
Testing all the modules together				x

6 Benefits of the research

According to the authors' knowledge, this is the first time that possibilities are given to users to querying movements and trajectories in a dataset of videos in a drawing manner.

The deliverables of the project will be the following :

- D1 : HMI module
- D2 : Tracking module
- D3 : Data module
- D4 : A priori model module
- D5 : Ontology module
- D6 : Final project report and presentation

7 Profile of the team

7.1 Leader

De Beul Dominique - PhD student
Working for more than 10 years in the telecommunication industry (Philips, Intel). He decided to return to school and obtained a MSc in computer science and management from the Engineering faculty of the University of Mons in Belgium. He is now a PhD student working in the domain of video segmentation and recognition.

7.2 Other researchers needed

Specialists in the field of video or image segmentation, human body recognition, large database, ontology, oriented object programming (C++), web design/programmer. Every interested people are welcome as well as MSc student as PhD student.

References

[1] Pedram Azad, Ales Ude, Tamim Asfour, and Gordon Cheng. Image-based markerless 3d human motion capture using multiple cues. *Neuroscience*, 2006.

[2] Eirik Fauske, Lars Moland Eliassen, and Rune Havnung Bakken. A comparison of learning based background subtraction techniques implemented in cuda. *Norwegian Artificial Intelligens Symposium NAIS*, pages 181–192, 2009.

[3] Nigel H. Goddard. Human activity recognition. In M. Shah and R. Jain, editors, *Motion-Based Recognition*, Computational Imaging and Vision Series. Kluwer Academic Publishers, 1997.

[4] Alberto Menache. *Understanding Motion Capture for Computer Animation and Video Games - Second edition.* Elsevier, 2011.

[5] Thomas B. Moeslund and Erik Granum. A survey of computer vision-based human motion capture. *Comput. Vis. Image Underst.*, 81:231–268, March 2001.

[6] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.*, 104:90–126, November 2006.

[7] Massimo Piccardi. Background subtraction techniques: a review. In *SMC (4)*, pages 3099–3104, 2004.

[8] Ronald Poppe. A survey on vision-based human action recognition. *Image Vision Comput.*, 28(6):976–990, 2010.

[9] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local svm approach. In *International Conference on Pattern Recognition*, 2004.

[10] Steffen Staab and Rudi Studer, editors. *Handbook on Ontologies.* International Handbooks on Information Systems. Springer, 2004.