Device based gesture recognition*

Damien Zufferey
Department of Informatics (DIUF)
University of Fribourg
Bd de Perolles 90
CH-1700 Fribourg
Switzerland
damien.zufferey@unifr.ch

ABSTRACT

Today, there is a growing interest in research and development of new human-machine interaction systems that are more natural and ergonomic for the users. The gesture recognition plays an important part of human-machine interaction systems. In this paper, we propose a survey on device based gesture recognition. The focus is done in systems that are based on accelerometers, on optical/magnetic tracking and on glove based equipments. First, we try to give a clear definition for posture and for (static/dynamic) gesture. Then, based on several papers, we describe the process for different types of approach and then we give an evaluation of them. Some applications of these technologies are also presented. Finally, as conclusion, we expose the oportunities and the issues related to these different approaches of gesture recognition.

Keywords

gesture, posture, recognition, device, accelerometer, glove

1. INTRODUCTION

With the advent of new technologies in the field of interactions between humans and machines, the focus is in finding new opportunities to reduce the barrier between them. Indeed interaction based on movements of the hand, arm, head, etc. is much more natural than interaction made through conventional devices like mouse or keyboard. This type of interaction is the heart of immersive virtual environments [1] such as the 3D desktop metaphor. These new ways of interaction are growing in commercial applications. The market offers powerful computers and multimodal devices that are available at most people in reasonable price.

However, these new interraction approaches are generally not very popular today, apart from in specific areas. This reflects the difficulty to recognize gestures and to interpret their meaning. Whatever the used device, we must deal with a quantity of data we received. Then, mathematical/statistical methods are needed for the gesture recognition process. Gesture recognition is a multistep process that varies depending on the used device and the type of gestures.

In this paper, we propose a survey on the state of the art concerning the gesture recognition. We are particularly interested in the gesture recognition through the use of a device which may be either a glove or an accelerometer-based hardware. The gesture recognition by video/image analysis is not addressed in detail in this paper.

This paper is ogranized as follow. First, we define some important concepts related to the gesture recognition. Secondly, we focus on the gesture recognition using a device based on accelerometers, including the Wiimote. Thirdly, we focus on the gesture recognition using a glove. Fourthly, we look briefly gesture recognition based on video/image analysis. Then finally, we summarize these different technologies and we discuss about advantages and issues of them.

2. GESTURE DEFINITION

First, we must give a definition for gesture. It is a difficult exercise to give a precise formal definition of what is a gesture. A gesture is a vague concept that has no single definition among the existing publications of gesture recognition. In [6], the authors remark that gestures are particularly related to the communicational aspect of the human hand and body movements. Webster dictionary, for example, defines gestures as *"the use of motions of the limbs or body as a means of expression; a movement usually of the body or limbs that expresses or emphasizes an idea, sentiment, or attitude"*.

In this document, we define a dynamic gesture as voluntary motions of hands and/or arms to express a clear action. This definition of dynamic gesture may well operate with devices based on accelerometers like the Wiimote. There is also static gesture (or posture). A static gesture is an orientation of hands and/or arms in the space during an amount of time, without any movements. For example, the hand of a person using a glove can express a particular static gesture. A dynamic gesture can be viewed as a sequence of static gestures. A gesture concerns not only the hand but can concern the whole body. However, in this document,

*This report was written as part of a master seminar about gesture recognition. More information is available on url: *http://diuf.unifr.ch/diva/web/site/index.php/teaching-seminars/10-seminars/125-gesture-recognition*

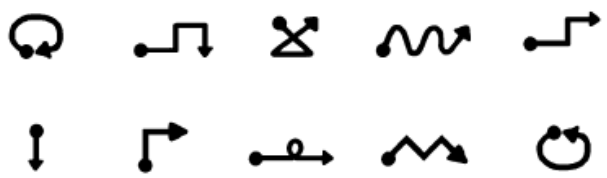

Figure 1: Sample of gestures to be recognized.

most of the time, we look at gestures performed by hands and/or arms. Note that a dynamic gesture is simply called "gesture" and a static gesture is called "posture".

3. ACCELEROMETER BASED

In this section we focus on devices that use accelerometers to measure movements. In [4] and [5], a SoapBox is used as accelerometer based device. In [7], a Wii Controller (Wiimote) is used as accelerometer based device.

3.1 Overview

In [7], the famous Wiimote from Nintendo is used for input of the user movements. This device is the main controller for the Wii console. An accelerometer in the controller is responsible to measure acceleration along three axes. An extension that contains a gyroscope can be added to the controller to improve rotation motions. The controller also contains an optical sensor allowing to determine where it is pointing. For that, a sensor bar highlighting IR LEDs is used. The connectivity of the controller is done via Bluetooth.

In the two others papers [4] and [5], a SoapBox is used. SoapBox is defined as Sensing, Operating and Activating Peripheral Box. This box is small and has low power consumption. It is equipped with 3-axis accelerometer, an illumination sensor, a electronic compass and a optical proximity sensor. For communication purpose, it is wireless with RF technologies. All the development can be written in C, utilizing the API offering.

In both papers, it is the 3-axis accelerometer component that was responsible to measure the movements.

The types of gestures to be recognized are for example: the trace of a square, a circle, a Z, ... In figure 1, we have some sample of gestures that comes from [4].

Possible applications of this type of gestures are rather wide. For example, [4] suggests to use gesture recognition to control home equipments such as VCR or TV.

3.2 Process

In all three papers, we have a system allowing the training and recognition of arbitrary gestures with the use of 3-axis accelerometers. With this type of device, we must deal with spatial and temporal data. We need a mathematical process to exploit these signals. Gestures are represented with data vectors representing the current acceleration of the controller in 3-axis. Theses vectors are analyzed to train and to recognize patterns for distinct gestures.

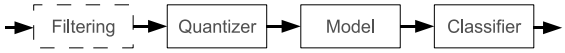

Figure 2: The process of gesture recognition with accelerometer based device.

The process consists of 3 phases as seen in figure 2. First we have the "Quantizer" that is responsible to cluster the data using a k-mean algorithm. Then, the "model" is a discrete Hidden Markov Model (HMM) that is used to train/recognize characteristic patterns for distinct gestures. Finally, a "Bayes-classifier" is used to select the appropriate gesture. Before these 3 phases, a filtering is applied on data for simplification purpose.

Filtering

In [7], two filters are applied to the vector data to get a minimum representation of a gesture. The first filter is a simple threshold eliminating all vectors which do not have a significant contribution to the characteristic of a gesture. The second filter is responsible to eliminate all vectors which are roughly equivalent to their previous.

In [4] and [5], the gesture data are first interpolated or extrapolated if the data sequence is too short or too long, respectively. Then, the amplitude of the data is scaled. These steps are improving the recognition of the gesture by the normalization of gesture speed variation.

Quantizer

Acceleration sensor produces to much data. Before putting them in a HMM, we need to cluster and abstract them. A k-means clustering method is used to partition the n observations into k clusters. Each observation belongs to the cluster with the nearest mean. The k is the number of clusters, their collection forms the codebook.

In [7], true 3D gestures are evaluated with a k of size 14 which was found empirically.

In [4] and [5], the size of the codebook k is 8 which was also found empirically. But, in constrast to [7], 3D data are converted to 1D vectors.

Model and Classifier

The gesture recognition system work in two phases: training and recognition. The training consists of several repetitions of each gesture that must be recognized later.

In [4], [5] and [7] a discrete Hidden Markov Model is used for the training and the recognition of gestures. A HMM is stochastic signal modeling method in which the system being modeled is assumed to be a Markov process with unobserved states. The global structure of the recognition system is composed with trained HMM for each gesture. The classification is done with a naive Bayes classifier that is a simple probilistic classifier.

3.3 Evaluation

In [7], an evaluation was conducted by collecting quantitative data to determine the percentage of correct recognitions. Several people representing different ages have been

selected. A serie of gestures to be recognized has been established. Each participant had to perform each gesture several times for training purpose. Then, again, he made each gesture one time for the recognition phase. The recognition rate reached 90 percent.

In [4], [5], the methodology for the evaluation was like in [7]. The goal is to obtain a recognition rate of over 95 percent, but by having the minimal number of required training repetitions in order to make the training process less laborious for the user. The recognition rate of over 95 percent was reached with a training consisting of 6 repetitions for each gesture.

4. GLOVE BASED

In [1] and [3] a gesture recognition based on gloves is described.

4.1 Overview

A data glove is a glove-like input device often used for virtual reality environments. It is equipped with various technologies such as a system for detection of bending of fingers. Often a motion tracker is attached to capture the global position/rotation data of the glove.

In [1], a P5 Glove from Essential reality was used. It is an inexpensive ($\backsim$ 50 Euro) glove with integrated 6 DOF tracking designed as a game controller. 6 DOF means six degrees of freedom, in fact the ability to move forward/backward, up/down, left/right (translation in three perpendicular axes) combined with rotation about three perpendicular axes (pitch, yaw, roll). The glove consists of five bend sensors to track the flexion of the wearer's fingers. An infrared-based optical tracking system is used to compute the glove position and orientation without the need for additional hardware. The glove is connected with a cable to the base station.

In [3], their gesture recognition system is based on two different components. First, two "CyberGlove" from Virtual Technologies are used, for each hand. This glove has flexible sensors that measure the position and movement of the fingers and wrist. Then, five "Flock of Birds" from Ascension Technology Corporation are used for six degrees-of-freedom tracking. This 6DOF tracking system sensor is based on magnetic technology. Note that the electro-magnetic field is distorted by metallic objects. Two are attached to the mounting point of each glove. Another one is mounted on a light-weight helmet worn by the user. The two others are attached to the subjects upper arms to register the position and orientation of the humerus.

4.2 Process

In [1], an important aspect is that a gesture is see as a sequence of successive postures. Postures in the recognition engine are composed of the flexion values of the fingers, the orienation data of the hand and an additional value to indicate the relevance of the orientation for the posture. These postures are taught to the system by simply performing them, then associating an identifier with the posture.

The recognition engine is divided into two components: the data acquisition and the gesture manager.

Data acquisition

The data acquisition component is responsible for processing the received data and then transmit them to the gesture manager. First, a set of filter is used to optimize the data. For example, the position/orientation information is very noisy due to dependance of lighting conditions. Thus, orientation data that exceed a given limit are discarded as improbable and replaced with their previous values. This type of filters are applied: deadband filter, dynamically adjusting average filter. Note that to be recognize as a posture, the user has to hold a position between 300 and 600 milliseconds in order to allow the system to detect a posture.

Gesture manager

The gesture manager is the principal part of the recognition system. This library maintains a list of known postures. The system try to match incoming data with existing posture. This is done by first looking for the best matching fingers constellation. Five dimensional vectors represent the bend values of the fingers and for each posture definition the distance to the current data is calculated. Then, the position/orientation data is compared in a likewise manner. Finally, in this gesture recognition system, a gesture is just a sequence of successive postures. For example, let's consider the detection of a "click" gesture. This gesture is defined as a pointing posture with outstretched index finger and thumb and the other fingers flexed, then a tapping posture with half-bent index finger.

In [3], the process for gesture recognition is different and more complex because the recognition concerns the whole upper-limbs including the head. A multi-level process that leads from the recognition of upper-limbs signals to symbols is described. The first-level symbols describe types of gestures/postures such as hand-shape or hand-orientation. An abstract body model is used for the derivation of signals to first level symbols. This model can describe the complete posture/gesture of the upper-body. As for others systems, data received from sensors are loaded with noise. Different types of filters must be applied to remove noise and to optimize the signals. Then, a second-level symbols are derived from the first-level symbols. This second level symbols constitue the application-specific semantic units. The goal of this approach is the possibility and the only necessity to adapt the second-level symbols according to the required interpretation of symbols by an application.

4.3 Evaluation

In [1], the evaluation is made with the help of a demo application. This application represents a virtual desktop with several gestural interaction possibilities. This desktop contains various type of objects such as documents, folders, widgets, ... The interaction is made with adapted gestures according to the object we want to manipulate. Several users tested the demonstrational environment with initial difficulties due to the trouble with the glove. But after a short time, most users were able to interact in natural way with this virtual desktop. Sometimes it was necessary to individually adapt the definition of some postures. Note that this

demo can run on a consumer grade computer which performs 3D graphic rendering and gesture recognition.

In [3], the second-level symbolic gesture definitions are the units which can be integrated with speech tokens to form multi-modal utterances.

5. VISION BASED

In vision based gesture recognition, we can see the camera as a device. Therefore, a short introduction to this subject is done.

According to [6] and [2], vision based gesture recognition is the discipline that consists from an image or a video sequence to recognize gestures or postures of a person for example. This vision based system can use only one camera, or several in the goal of computing a 3D model of the scene.

Vision based gesture recognition is a multi-steps process that consists of:

Pre-processing

Usually before that an image can be exploited for gesture recognition, it is necessary to apply some filters or others transformations in order to have an image that satisfies certain assumptions implied by future used methods.

Segmentation

This phase consists of extracting from an image, image points or regions which are relevant for further processing.

Feature extraction

In this step, the goal is to extract features like contours, fingertips, ... These features are then used as characteristic patterns for the recognition engine.

Model/Classifier

This is the place of the recognition procedure. Often modeling tools like Hidden Markov Model are used for gesture training and recognition. Finally a classification algorithm is used.

The great advantage of vision based gesture recognition is the liberty that is given to the person in which gesture recognition is performed. As opposite to the device based gesture recognition that use gloves or accelerometers boxes, the vision based gesture recognition is a non-intrusive method. However, it can nevertheless be observed several disadvantages. Often, more computation power is required for vision based gesture recognition compared to device based gesture recognition. An other problem is the difficulty to perform vision-based gesture recognition when we have different backgrounds or different lighting conditions.

6. DISCUSSIONS

In this section, first I give my impression about the results of these different approaches for gesture recognition. Then, I give some observations about the papers that were used for writing this article.

For the part that concerns accelerometer based gesture recognition, I observed that recognition rate was good with an average of 90-95 percent. But I deplore that to many training were necessary to obtain a correct result, particular for [7]. Then, I remark that the collection of gestures that have to be recognized was often small, and gestures seem to be clearly different. What would be the result with a large collection of similary gestures ? Concerning glove based gesture recognition, I think it is an intrusive and laborious system for the user. This is particulary the case for [3].

For the three papers about accelerometer, I find that the process and the evaluation is clearly understandable. By opposition, the paper [3] which describes a recognition system: "from signals to symbols", is abstract and difficult to understand. In addition, a sophisticated concept is presented, but the evaluation of this system is not very clear.

7. CONCLUSIONS

A survey was presented about device based gesture recognition, in particular on devices equipped with 3-axis accelerometer, or based on gloves. With reasonable computing power, it is possible to obtain an acceptable recognition rate ($\sim$ 90 %) for the recognition of unsophisticated gestures. These results seem to be sufficient for most common applications. However under the condition of having enough "trained" the system, which can be laborious for the user. Accelerometer based devices seem to be less intrusive than gloves, but in constrast gloves can recognize postures and flexions of fingers.

8. REFERENCES

[1] M. Deller, A. Ebert, M. Bender, and H. Hagen. Flexible gesture recognition for immersive virtual environments. In *Tenth International Conference on Information Visualization (IV 2006)*, pages 563–568. IEEE, July 2006.

[2] W. Du and H. Li. Vision based gesture recognition system with single camera. In *5th International Conference on Signal Processing Proceedings*. IEEE, 2000.

[3] M. Froehlich and I. Wachsmuth. Gesture recognition of the upper limbs - from signal to symbol. *I. Wachsmuth and M. Froehlich (eds.): Gesture and Sign Language in Human-Computer Interaction. Berlin: Springer-Verlag (LNAI 1371)*, pages 173–184, 1998.

[4] J. Kela, P. Korpipaa, J. Maentyjaervi, S. Kallio, G. Savino, L. Jozzo, and S. Marca. Accelerometer-based gesture control for a design environment. *Springer. Personal And Ubiquitous Computing*, 10(5):285–299, 2006.

[5] J. Maentyjaervi, J. Kela, P. Korpipaa, and S. Kallio. Enabling fast and effortless customisation in accelerometer based gesture interaction. In *Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia*, pages 25–31. ACM, 2006.

[6] V. Pavlovic, R. Sharma, and T. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):677–695, 1997.

[7] T. Schloemer, B. Poppinga, N. Henze, and S. Boll. Gesture recognition with a wii controller. In *Proceedings of the Second International Conference on Tangible and Embedded Interaction (TEI'08)*. ACM, 2008.