J. Vis. Commun. Image R. 28 (2015) 97–104

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A hand gesture recognition technique for human–computer interaction ☆

Nurettin Çağrı Kılıboz, Uğur Güdükbay *

Department of Computer Engineering, Bilkent University, 06800 Ankara, Turkey

ARTICLE INFO

Article history:
Received 3 April 2014
Accepted 21 January 2015
Available online 2 February 2015

Keywords:
Dynamic gesture recognition
Hand gesture
Finite state machine-based recognition
Gestural interfaces
Gesture-based interaction
Human-computer interaction
Intuitive interfaces
Hand trajectory recognition
Adaptive gestures

ABSTRACT

We propose an approach to recognize trajectory-based dynamic hand gestures in real time for human–computer interaction (HCI). We also introduce a fast learning mechanism that does not require extensive training data to teach gestures to the system. We use a six-degrees-of-freedom position tracker to collect trajectory data and represent gestures as an ordered sequence of directional movements in 2D. In the learning phase, sample gesture data is filtered and processed to create gesture recognizers, which are basically finite-state machine sequence recognizers. We achieve online gesture recognition by these recognizers without needing to specify gesture start and end positions. The results of the conducted user study show that the proposed method is very promising in terms of gesture detection and recognition performance (73% accuracy) in a stream of motion. Additionally, the assessment of the user attitude survey denotes that the gestural interface is very useful and satisfactory. One of the novel parts of the proposed approach is that it gives users the freedom to create gesture commands according to their preferences for selected tasks. Thus, the presented gesture recognition approach makes the HCI process more intuitive and user specific.

1. Introduction

Various approaches to human–computer interaction (HCI) have been proposed in the last few decades as an alternative to the classic input devices of keyboard and mouse. However, these new techniques have not been able to supersede the old ones due to their lack of intuitiveness. Recently, HCI has regained popularity due to the intuitive and successful interaction techniques of devices such as tablet PCs, smart phones and even smart houses. All these applications use voice commands, mimics, and gestures to interact with humans.

Human–computer interaction with hand gestures plays a significant role in these modalities because humans often rely on their hands in communication or to interact with their environment. Therefore, hand-gesture-based methods stand out from other approaches by providing a natural way of interaction and communication [1]. Many studies evaluate gesture-based interaction techniques, their drawbacks, and propose ways to increase their effectiveness [2–4].

There exist various definitions of hand gestures in the literature. Some studies define gestures as only static postures [5], while others consider hand motions and trajectory information as a part of the gestures [6]. In the scope of this study, we consider only the motion trajectory of the hand (excluding finger bending and orientation information) to define gestures.

Recognizing gestures is a comprehensive task combining various aspects of computer science, such as motion modeling, motion analysis, pattern recognition and machine learning [7]. Since the beginning of the 1990s, many hand gesture recognition techniques have been proposed. These studies can be divided into two categories, based on their motion capture mechanism: *vision-based* or *glove-based*. Vision-based techniques rely on image processing algorithms to extract motion trajectory and posture information [8–10]. Their success highly depends on the used image analysis approaches, which are sensitive to the environmental factors, such as illumination changes, and may lose fine details due to hand and finger occlusion [11].

Glove-based techniques generally provide more reliable motion data and eliminate the need for middle-tier software to capture hand positions and postures [12]. On the other hand, they require the user to wear cumbersome data gloves and position trackers, and usually carry a few connection cables. These factors reduce intuitiveness and usefulness of these methods and make them costly [12].

Recent developments in technology pave the way for more accurate and affordable motion capture technologies, namely

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author. Fax: +90 (312) 266 4047.
E-mail address: gudukbay@cs.bilkent.edu.tr (U. Güdükbay).

http://dx.doi.org/10.1016/j.jvcir.2015.01.015

depth camera sensors such as Kinect™ and Wii™. Hence, it is possible to retrieve more precise motion data without the limitations of traditional (vision-based and glove-based) approaches. Even small objects like human fingers can be effectively captured by these devices [13]. Researchers also propose dynamic hand gesture recognition algorithms that utilize these devices [14]. Similar to ours, these studies use hand trajectory based gesture recognition algorithms. Although these studies claim that their recognition rate is over 90% with relatively simple and small gesture sets, they use fairly large training sets. Unlike our approach, none of these approaches are capable of recognizing gestures on the fly. Additionally, our gesture recognizer does not require large training sets.

Studies in this field can also be classified by examining whether they recognize static or dynamic gestures. Although static gesture recognition is relatively simpler, it still requires much effort due to the complexity of gesture recognition. Most of the static gesture recognition research focuses on neural-network-centered approaches [15,16], but for dynamic gesture recognition, hidden Markov model (HMM)-based approaches are generally preferred because they yield better results [17–19]. Similar to our study, finite-state machine (FSM)-based techniques [20–22] are also used to recognize dynamic gestures. Other studies suggest using fuzzy logic [23] and Kalman filtering [24] for gesture recognition.

Many gesture recognition techniques such as neural-network [25] and HMM-based [26] approaches require a preliminary training phase in which an extensive training data is fed the system to form the recognizers. Our approach can achieve similar recognition rates without requiring a large training set or on-line training. The other advantage that we obtain from the FSM-based recognizer is that we can detect gestures in a stream of hand motion unlike the other methods [27] where the start and end positions of gestures should be specified explicitly.

We introduce an intuitive approach to teach a machine to recognize a hand gesture command so users can apply them to devices such as TVs or eReaders. This approach allows users to create their own gesture commands for a particular task according to how they think it suits the action.

Similar to the other techniques, the proposed approach consists of two stages: *learning* and *recognition*. In the learning stage, the user is asked to repeatedly perform a particular gesture. The system records the motion trajectory of each gesture sample with a magnetic 3D position tracker attached to the user's hand. Unlike the other approaches [28], motion data is collected in gradient-like form. Instead of noting the absolute position of the hand, its position relative to the previous recording is noted. Additionally, threshold-based filtering is applied to the collected data to reduce noise caused by unintended vibrations and hardware errors. Next, collected motion data is filtered using a component-based sliding window technique for smoothing and further noise removal. Then, the filtered trajectory information is transformed into our gesture representation format, which is basically an ordered sequence of events (directional movements).

In the last step of the learning phase, our approach chooses a few event sequences (using the Needleman–Wunsch sequence-matching algorithm [29]) from the provided samples to form a base for gesture recognizers. The algorithm compares every pair of event sequences (gesture pairs) and computes a similarity score for them. The event sequences with the highest similarity scores are selected to form the bases for the gesture recognizers. Then, a recognizer finite state machine (FSM) is generated based on these chosen gestures. Because FSMs are sequence recognizers, each forward transition in a generated FSM corresponds to an event in the selected sequence in the respective order. This learning phase is repeated for every distinct gesture, with several FSMs produced for each.

In the recognition stage, continuous inputs from the tracker are processed in a similar manner as in the learning stage and fed to all the recognizer machines. If one of the previously captured event sequences occurs during the session, the respective recognizer machine traverses all the states and reach the final (accepting) state. The resulting gesture recognition event triggers the action assigned for the gesture. With this approach, gestures can be recognized in real time.

One important feature of the proposed dynamic gesture recognition technique is that it can effectively detect gestures in a motion flow regardless of the motion capture technique. Vision-based approaches can be used in the proposed gesture recognition framework instead of the glove-based hand motion capture. The proposed gesture representation and recognition mechanism is especially suitable for vision-based hardware and algorithms. In fact, vision-based approaches may overcome the major problems of the hardware used because they will address the limitations of the device such as restricted motion capture range and carrying an uncomfortable attachment. For example, the results of hand follower algorithms proposed in [30,31] can be easily converted to our gesture representation and can be fed to the recognizer machines. It is even possible to extend the usage area of our approach to the public spaces using the hand segmentation and recognition approaches described in [32] that generate hand coordinates, which is sufficient for us to recognize hand gestures.

The rest of the paper is organized as follows: The proposed approach is described in detail in Section 2. The details of the conducted user study and experimental results presented in Section 3. Analysis and discussion on the experimental results are given in Section 4. Section 5 provides conclusions and future work.

2. Proposed approach

2.1. Gesture representation

In gesture recognition, representing gestures is a critical issue. We define gestures as a series of events performed consecutively. For trajectory-based dynamic gestures, this is a valid definition because trajectories are a series of directional vectors combined in a particular time interval. In our case, events are directional movements and a gesture is an ordered sequence of these directional movements (see Fig. 1).

In this study, we limit the trajectories to the *xy*-plane for simplicity. Our representation not only allows creating many interesting gestures, it also improves the robustness of the algorithm. It is possible to extend the event (gesture) alphabet with the third dimension, or with other features such as finger movements. Using

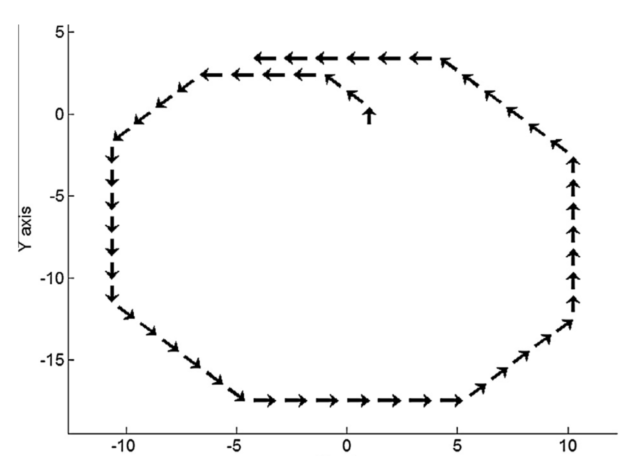

Fig. 1. A gesture (circle) is represented as an ordered sequence of directional movements.