ARTICLE IN PRESS
JID: YCVIU [m5G;October 27, 2016;15:30]

Computer Vision and Image Understanding 000 (2016) 1–11

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

A dynamic gesture recognition and prediction system using the convexity approach

Pablo Barros[a,*], Nestor T. Maciel-Junior[b], Bruno J.T. Fernandes[b], Byron L.D. Bezerra[b], Sergio M.M. Fernandes[b]

[a] Department Informatics, University of Hamburg, Knowledge Technology, Hamburg, Germany
[b] Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife-PE, Brazil

ARTICLE INFO

Article history:
Received 14 September 2015
Revised 14 September 2016
Accepted 17 October 2016
Available online xxx

Keywords:
Gesture recognition
Computer vision
Features extraction
Gesture prediction

ABSTRACT

Several researchers around the world have studied gesture recognition, but most of the recent techniques fall in the curse of dimensionality and are not useful in real time environment. This study proposes a system for dynamic gesture recognition and prediction using an innovative feature extraction technique, called the Convexity Approach. The proposed method generates a smaller feature vector to describe the hand shape with a minimal amount of data. For dynamic gesture recognition and prediction, the system implements two independent modules based on Hidden Markov Models and Dynamic Time Warping. Two experiments, one for gesture recognition and another for prediction, are executed in two different datasets, the RPPDI Dynamic Gestures Dataset and the Cambridge Hand Data, and the results are showed and discussed.

1. Introduction

Gesture recognition systems can be used as a natural and welcoming interface of interaction between computational systems and humans. These systems use human movement patterns to identify, learn, and generalize gestures executed by a user. There are several applications in the area of gesture recognition, such as games (Lee and Hong, 2010; Rautaray and Agrawal, 2011), human-robot interaction (Bodiroza et al., 2013; Lee, 2006), interaction with televisions (Jeong et al., 2012), and sign language recognition (Ciaramello and Hemami, 2011; Liu and Xiao, 2015; Silanon and Suvonvorn, 2014; Zhou et al., 2008).

Nowadays, it is possible to capture gestures executed by humans using only a video camera on a smartphone, tablet, or notebook. Since most people have at least one of these devices in their possession, using gestures for communication with computational systems could be employed more often than in the previous years. The evolution of computers, in hardware and software, also increases the usage of gestures as an important tool for human-computer communication.

Gesture recognition systems can be clustered into three different categories (Ibraheem and Khan, 2012): systems based on gloves or external sensors attached to a user for gesture capture (Cheng et al., 2015; Dekate et al., 2014; Han, 2010; Huang et al., 2011; Jeong et al., 2011; Xinyu et al., 2010), systems that recognize a gesture through a tracking device, such as a mouse (Bhattacharjee et al., 2015; Chivers and Rodgers, 2011; Cho et al., 2004; Jeong et al., 2012; Schlecht et al., 2011), and systems that capture gestures using a video camera and process them with computer vision techniques (Bernardes et al., 2009; Koceski and Koceska, 2010; Leubner et al., 2001; Sen et al., 2005; Wu et al., 2015; Zhang and Zhang, 2008). The first category captures the gesture more precisely, but the process is invasive, as the user wears the sensors around him. The application of gloves for capturing happens in controlled environments, where the gloves are connected to computers, but the capturing process becomes complicated to be used in an external environment and real world applications.

The second category of gesture recognition systems uses a tracking technique to follow a hardware device in the screen. The tracked path is the gesture to be recognized. This kind of recognition uses a simple gesture definition, reducing the computational cost. However, the possible gestures represented are less significant and less precise than the other categories.

The last category uses a video camera to capture and identify the gesture. The recognition process uses the images to extract some features, such as movement, position, velocity, color,

* Corresponding author.
E-mail addresses: barros@informatik.uni-hamburg.de (P. Barros), ntmj@ecomp.poli.br (N.T. Maciel-Junior), bjtf@ecomp.poli.br (B.J.T. Fernandes), byronleite@ecomp.poli.br (B.L.D. Bezerra), smmf@ecomp.poli.br (S.M.M. Fernandes).

http://dx.doi.org/10.1016/j.cviu.2016.10.006
1077-3142/

Please cite this article as: P. Barros et al., A dynamic gesture recognition and prediction system using the convexity approach, Computer Vision and Image Understanding (2016), http://dx.doi.org/10.1016/j.cviu.2016.10.006

among others. This approach can describe a more complex gesture than the second category, and it does not need to be used in a controlled environment. This category uses computer vision techniques that can be embedded in a smart-phone, for example, and used for gesture recognition in any place.

One of the main problems with dynamic gesture classification is to deal with sequence recognition. Gestures are naturally dynamic, as a sequence of movements and postures, and each step has influence in further ones. These works deal usually with temporal dependency (Alon et al., 2009; Chang, 2016; Santos et al., 2015), and usually demand very long time and computational effort for training. Besides that, a robust feature extractor is necessary, which increases the computational cost of the process (Frolova et al., 2013; Wu et al., 2016), and thus decreases the possibility of the model to work in real time.

A desirable characteristic in a gesture recognition system is the ability to recognize gestures in real time. As pointed out by Mori et al. (2006), a real time gesture recognition system has to be able to predict the gesture that is being executed before it ends. Prediction allows the recognizer to work in real time and makes the classifier more accurate, as it can use the prediction results to improve recognition. The prediction attempts to identify a pattern that has yet to be completed. Some prediction techniques have been used with success to improve speech recognition (Helander and Nurminen, 2007; Husssain et al., 2009; Javed and Ahmad, 2014; Satya et al., 2011; Stavrakoudis and Theocharis, 2007). A few studies use the prediction concept gesture prediction, as described by Ahmad et al. (2015); Kohlsdorf et al. (2011); Liu and Xiao (2015); Silanon and Suvonvorn (2014).

Although there are many studies using computer vision for gesture recognition, some problems remain, like significant computational and time costs for the algorithms. To use gesture recognition techniques in a real time environment, it is necessary to reduce their computational and time costs. Such reduction can be achieved using a smaller feature vector to describe a gesture or a prediction technique (Hasan and Kareem, 2012). Our study shows a dynamic gesture recognition system that uses an innovative technique, called the Convexity Approach, for feature extraction. The system is evaluated for dynamic gesture recognition and for gesture prediction. It implements classification and prediction modules based on Hidden Markov Models and Dynamic Time Warping. The proposed method is able to classify dynamic hand gestures by recognizing individual hand postures and modeling a sequence with them. This way, our approach can classify gestures with different speed and different users. Our approach is strongly based on an innovative feature extraction technique, and it uses the capability of dynamic time warpers to model gesture sequences.

This paper is organized as follows. Section 2 describes the Convexity Approach technique. Section 3 shows the recognition and prediction system architecture. Section 4 presents the experimental setup and results, as well as a discussion about the obtained results. Finally, in Section 5, we present the concluding remarks.

2. Convexity Approach

Several techniques are used to describe gestures. Most of these feature extraction techniques are affected by the curse of dimensionality (Bilal et al., 2011). The curse of dimensionality states that an approximation of a numerical function will have a higher computational cost if its variables increase (Kouiroukidis and Evangelidis, 2011). There are proposed solutions for the curse of dimensionality, such as the reduction of the feature vector dimension (Pagel et al., 2000; Zhao et al., 2010), classification algorithm optimization (Qaiyumi and Mirikitani, 2006; Qin and Tang, 2009), and the use of feature selection strategies for the problem (Teoh and Sheble, 2007).

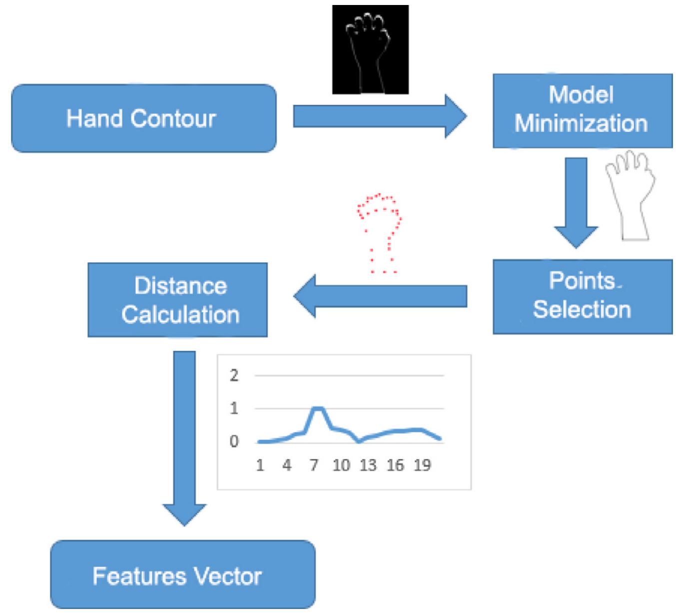

Fig. 1. Illustration of Convexity Approach execution.

The Convexity Approach, the feature extraction technique, can describe a hand gesture using only dynamically selected points in the hand contour. The selection of points is minimized for each hand posture, so the feature vector contains the minimal features necessary to describe the hand.

The Convexity Approach extracts features of one image at a time. This image must contain only the hand contour. The first step of the algorithm is to reduce the geometrical model of the hand, eliminating curves. The second step is to find a minimal set of points that can represent the minimized hand model. The last step is to extract the distance of these points and create a feature vector that will describe the hand. Fig. 1 shows the illustration of the Convexity Approach execution.

2.1. Model minimization

The first step of the Convexity Approach assures that any extra information will not be extracted. In the hand gesture context, extra information is a part of the hand that can be excluded without losing the shape of the geometrical model of the hand. For example, a curve in the hand can be represented by three points. The Douglas-Peucker (Douglas and Peucker, 1973) algorithm is used to create the minimized hand model and it is successfully applied for traffic sign recognition (Soendoro and Supriana, 2011), model minimization using compression (Nandakumar et al., 2005), and geographical applications (Youfu and Guoan, 2010).

An ordered set of $n+1$ points in a plane forms a polygonal chain. Given the chain C with n segments, the Douglas-Peucker algorithm will generate a chain C' with fewer segments than C. The two endpoints of a set of points are connected by a straight line $\overline{AB}$ as the first rough approximation of the polygon. Iterating over all the vertexes, v, of all segments n in C, the distance between the vertex v_i and the center of $\overline{AB}$ is calculated. If the distance of all vertexes is shorter than a threshold t, then the approximation is good, the endpoints are retained, and the segment $\overline{AB}$ will be added to C' and represent the polygon. However, if any of these distances exceeds t, the approximation is not good. In this case, it chooses the furthest point P, and subdivides the initial set points into two new segments $\overline{AP}$ and $\overline{PB}$. The same procedure is repeated recursively on these two new segments, and the new segments are