Signal Processing: *Image Communication* 30 (2015) 178–189

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Stratified gesture recognition using the normalized longest common subsequence with rough sets

Clementine Nyirarugira, TaeYong Kim *

Chung-Ang University, Seoul, South Korea

ARTICLE INFO

Article history:
Received 14 April 2014
Received in revised form
21 October 2014
Accepted 23 October 2014
Available online 4 November 2014

Keywords:
Gesture vocabulary
Rough set theory
Gesture recognition
Segment based longest common subsequence
Dynamic hand gesture recognition

ABSTRACT

In this paper, we propose a stratified gesture recognition method that integrates rough set theory with the longest common subsequence method to classify free-air gestures, for natural human–computer interaction. Gesture vocabularies are often composed of gestures that are highly correlated or comprise gestures that are a proper part of others. This reduces the accuracy of most classifiers if no further actions are taken. In this paper, gestures are encoded in orientation segments which facilitate their analysis and reduce the processing time. To improve the accuracy of gesture recognition on ambiguous gestures, we generate rough set decision tables conditioned on the longest common subsequences; the decision tables store discriminative information on ambiguous gestures. We efficiently perform stratified gesture recognition in two steps: first a gesture is classified in its equivalence class, under a predefined rough set indiscernibility, and then it is recognized using the normalized longest common subsequence paired with rough set decision tables. Experimental results show an improvement of the recognition rate of the longest common subsequence; on preisolated gestures, we achieve an improvement of 6.06% and 15.09%, and on stream gestures 19.79% and 28.4% on digit and alphabet gesture vocabularies, respectively.

1. Introduction

Since the early days of computers, human–computer interaction (HCI) has evolved considerably, from the use of keyboards and mice, through the use of touch screens, to natural user interaction interfaces using gestures [1]. Gesture recognition systems that allow free-air movements has been recently introduced into the home entertainment environment such as smart TV control and gaming [2]. Motion and vision are the two primary sensors used in natural HCI technologies. Vision based gesture recognition systems offer a noncumbersome and intuitive way of interaction compared to wearable motion sensor based gesture recognition systems.

In vision based HCI, a camera device is used to acquire data. A gesturing object such as a hand is then segmented from the rest of the image, and its centroid is considered as the hand position. A sequence of hand positions constitutes a gesture trajectory. Dynamic hand gestures encode information by temporal trajectories that may be mapped into descriptive features. The resulting features are then fed into a gesture recognition method which decide whether the trajectory is a specific gesture in a gesture vocabulary or a nongesture. When a gesture is recognized, its corresponding description is transformed into a command that is sent into a game [2] or TV set to trigger some action or reaction.

One of the key issues in HCI interfaces or control systems is the precision of the gesture recognition methods. One of the challenges is that gestures vocabularies often contain highly correlated gestures that lead to ambiguous recognition. Among highly correlated gestures, we separate subgestures, i.e., gestures that are proper parts (completely contained) of

* Corresponding author. Tel.: +82 2 812 5717.
E-mail addresses: rugiracle@cau.ac.kr (C. Nyirarugira), kimty@cau.ac.kr (T. Kim).

http://dx.doi.org/10.1016/j.image.2014.10.008
0923-5965/

others, from gestures that have a big overlapping part. These kinds of gestures are ambiguous and can lead to gesture substitution or multiple recognition, this will decrease the degree of precision of recognition algorithms, if no further actions are taken. It is important to investigate internal gesture vocabulary relationship and identify which gesture categories tend to be confused in order to take adequate measures.

Subgestures identification and reasoning problem is a challenging task. A few attempts to address the problem of subgestures have been made. In [3], a two-layers conditional random field (CRF) architecture has been used to deal with the problem of signs that appear as a part of other signs. The second CRF layer is used as a subgesture reasoning layer. However, CRF based recognition methods are complex and can be computationally expensive as stated in [4]. In [5], subgesture relationships are manually specified and rules are derived to correctly recognize supergestures. Furthermore, a few simple rules coupled with learned subgestures show an improvement in recognition accuracy of digit gestures [6]. However, a general intergesture category competition method is still needed.

We choose the longest common subsequence (LCS) as the base of gesture vocabulary analysis. The LCS is a similarity measure that has been used successfully for string matching. It is a global alignment method that is robust to noise and outlier [7]. The LCS based recognition method has been used successfully in dynamic hand gesture recognition systems [4,8]. These methods outperform the hidden Markov model (HMM) based gesture recognition method, which is considered as the state of the art. Although, the methods in [4,8] show promising results, a subgesture reasoning method is still needed. The LCS does not reveal the difference between the sequences that are to be matched. As a drawback, LCS based gesture recognition methods are challenged by ambiguous gestures.

In this paper, we propose a solution to the problem of subgestures by integrating rough set theory (RST) to the LCS. We use RST to learn the discriminative information among gesture classes. RST is a mathematical approach to imperfect knowledge [9]. It deals with vagueness and uncertainty. It has been used in many areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, decision support system, inductive reasoning and pattern recognition [10].

Within the gesture recognition context, rough set represents inter-gesture category relationships through the use of approximations. Each gesture category is defined in terms of its upper and lower approximations. Approximations are generated using a set of gesture attributes/features. The vagueness or ambiguity of a gesture category is expressed by a boundary region between the approximations. A boundary region between two gesture categories comprise gestures that can be classified in either category, under a set of attributes. The boundary region is inversely proportional to the accuracy of the approximation. We focus on gesture categories with a nonempty boundary. A set of discriminating attributes between two ambiguous gesture categories is then generated conditioned on their LCS. Condition attributes group ambiguous gesture categories in an equivalence class and discriminative attributes are identified inside these equivalence classes.

A gesture input is recognized in two steps; first, it is classified into its equivalence class, under a predefined set of attributes. Then, a final classification is done inside the equivalence class using the normalized LCS. The two step stratified classification improves the recognition rate and reduces considerably the number of templates to be matched. We evaluate the proposed method on character input formed by two gesture vocabularies; which are English alphabet and digit number. Both vocabularies comprise highly correlated gesture and subgesture relationships. The proposed method shows promising results. The remainder of this paper is organized as follows. Section 2 contains an overview of relevant research in the field of dynamic gesture recognition. Section 3 presents the architecture of the segment based normalized longest common subsequence paired with RST, Section 4 shows experiments performed to evaluate the proposed method, and finally, Section 5 concludes the work.

2. Related work

Dynamic hand gesture recognition is a difficult and challenging problem that has been addressed in many ways. A widely used approach is the hidden Markov model [11,12]. HMM based gesture recognition methods represent each gesture by a set of states associated with probabilities (initial, transition, and observation) learned from the training examples. HMM recognizers choose a model with the best likelihood and classify a given gesture to the corresponding gesture category. Although HMM recognition systems choose a model with the best likelihood, it is not guaranteed that the pattern is really similar to the reference gesture unless the likelihood value is high enough, above some threshold. In the case a simple threshold does not work well, a sophisticated threshold model can be derived [13,14].

CRFs have been used in gesture recognition systems [3,11]. CRFs offer several advantages over hidden Markov models, including the ability to relax the strong independence assumption made in HMM models. Sign language spotting with a threshold model based on CRF has been proposed [3]. A method for designing a threshold in the CRF model was proposed which performs an adaptive threshold for distinguishing between sign in vocabulary and nonsign patterns. A short sign detector and subsign reasoning method are included to further improve sign language spotting accuracy. A subsign reasoning method is employed to avoid premature detection of a sign that shares movements with other signs. During classification, the first layer uses the threshold model with CRF to discriminate in vocabulary signs and nonsign patterns. In the second layer subsign patterns are recognized in the spotted sign sequence by the first layer. HMM and CRF based gesture recognition methods have comparative performance [15], in general. The main disadvantages of HMM and CRFs are that they require a large number of samples and long training time to calibrate the models [8].

Dynamic time warping (DTW) is another approach often used for dynamic gesture recognition task [16–18].