Pattern Recognition Letters 000 (2017) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

# Continuous hand gesture recognition based on trajectory shape information

Cheoljong Yang[a], David K. Han[b], Hanseok Ko[a,*]

[a] Department of Visual Information Processing, Korea University, Seoul, Republic of Korea
[b] Office of Naval Research, Arlington, VA, USA

ARTICLE INFO

*Article history:*
Available online xxx

*Keywords:*
Gesture recognition
Human robot interaction
Convolution neural network
Conditional random fields
Trajectory segmentation
Trajectory shape modeling

ABSTRACT

In this paper, we propose a continuous hand gesture recognition method based on trajectory shape information. A key issue in recognizing continuous gestures is that performance of conventional recognition algorithms may be lowered by such factors as, unknown start and end points of a gesture or variations in gesture duration. These issues become particularly difficult for those methods that rely on temporal information. To alleviate the issues of continuous gesture recognition, we propose a framework that simultaneously performs both segmentation and recognition. Each component of the framework applies shape-based information to ensure robust performance for gestures with large temporal variation. A gesture trajectory is divided by a set of key frames by thresholding its tangential angular change. Variable-sized trajectory segments are then generated using the selected key frames. For recognition, these trajectory segments are examined to determine whether the segment belongs to a class among intended gestures or a non-gesture class based on fusion of shape information and temporal features. In order to assess performance, the proposed algorithm was evaluated with a database of digit hand gestures. The experimental results indicate that the proposed algorithm has a high recognition rate while maintaining its performance in the presence of continuous gestures.



## 1. Introduction

Recently, demand for intuitive and effective interfaces between humans and robots has emerged as a response to the rapidly increasing use of robots in various industries. In the field of human-robot interaction (HRI), significant efforts have been made to develop speech and gesture recognition [18,24,28]. Although speech recognition has numerous applications due to its consistent performance in handling natural expressions, issues such as the influence of noise or the distance between a microphone and its user remain as major limitations. To overcome these limitations, gesture-based HRI has started to gain research interest. Hand gesture based methods are capable of providing robust performance regardless of distance or background noise. Therefore, gesture recognition can be considered as one of the key modalities of intuitive HRI systems.

In gesture-based HRI, hand trajectory information, such as position or direction, is a key component of the communication between the user and the robot. To extract this information, contactless sensors have been introduced to replace highly inconvenient gesture input devices [15,21]. Although an interface based on contactless sensors does not require user-borne hardware, its camera typically receives continuous hand motion information even when hand motions are unintended. Therefore, in contactless environments, a recognition system has to be able to extract and classify intended trajectories only. Some key issues that affect contactless hand gesture recognition include difficulty in clearly establishing the location of a gesture is a continuous stream of data, variation in the speed of gestures between individuals, and variation in gesture trajectories between individuals [1,16]. In the first case, a recognition system requires a method to accurately determine when an intended gesture occurs in a given continuous motion stream. For a certain set of gestures, such as the writing of alpha-numeric characters, the beginning and the ending of each symbol are often uncertain due to “garbage motions” connecting two consecutive symbols. This is a serious difficulty for most contactless hand gesture recognition systems that employ a sliding window of fixed width. Variations in gesture speed may also result similar difficulties. As in the case of handwriting, variation in the trajectories of hand gestures occur between individuals and these differences in trajectory may adversely affect recognition performance. Therefore, practical HRI systems must be capable of robust segmentation and recognition to handle unknown gesture duration and variation between individuals.

There are two major approaches to the segmentation and recognition of hand gestures: 1) segmentation prior to recognition,

* Corresponding author.
*E-mail addresses:* hsko@korea.ac.kr, hanseok.ko.34@gmail.com (H. Ko).

http://dx.doi.org/10.1016/j.patrec.2017.05.016
0167-8655/

and 2) simultaneous segmentation and recognition. The former approach requires an extra set of gestures, such as cue motions or pauses, between gestures to ensure effective performance [31,36]. The latter approach does not require extra gestures, thus it is considered to be more natural [10,19,32]. As such, we focus on the second approach, treating segmentation and recognition as a simultaneous process.

In simultaneous segmentation and recognition, the main objective is to determine the optimal gesture segment in the temporal domain which provides accurate gesture matching results.

Most simultaneous processes are based on sequential feature modeling developed for isolated gestures. Representative isolated gesture modeling methods are Hidden Markov Model (HMM) [2,5], Conditional Random Fields (CRFs) [27,33] and Dynamic Time Wrapping (DTW) [7,26].

For practical application, it is necessary to expand these methods to include continuous gesture recognition. HMM and CRFs utilize probabilistic threshold models that establish adaptive criteria for distinguishing between gestures and non-gestures. Lee et al. proposed a non-gesture model, i.e. garbage motion, based on HMM [16]. The non-gesture model provides a confidence measure that is used as an adaptive threshold to determine the starting and end points of input patterns. Similarly, Yang et al. introduced an adaptive thresholding method based on CRFs to recognize signs and non-sign patterns [32]. Morency upgraded CRFs and incorporated hidden state variables which model the sub-structure of a class sequence and represent the dynamics between class labels [19]. Song proposed a method to recognize continuous and unbounded gesture by applying multi-layer LDCRF [23]. Along with statistical theoretical models, dynamic template matching, which searches for matches with pre-defined gestures, has also been considered for use in the recognition of gesture segments. Alon et al. [1] proposed a dynamic programming method with sub-gesture detection and reasoning which avoids premature gesture spotting. Of the dynamic template matching methods, DTW, which attempts to line up given sequences to gesture templates, is also a popular approach [9,17]. Although DTW methods have the advantage of low computational complexity, their sensitivity to noise and outliers degrades their performance. In an effort to improve the performance of DTW, an analysis of similarity measures between given sequences and gesture templates has been conducted [29]. Longest Common Subsequence (LCS), which ignores pairs that match poorly, is a popular measure of similarity [12]. By discarding poorly matched points, LCS provides a higher robustness to noise and outliers than does DTW. Stern et al. [25] proposed an extension of LCS, which produces candidate segments of a gesture trajectory while considering variations in gesture duration, based on detecting the change in feature state transition in near time.

In summary, both statistical and template matching methods perform recognition of gesture streams based on the state transition of temporal features. While conventional methods concentrate on the adjacent time characteristics of gesture trajectories, there may be value in focusing on their shape. The feature state transition methods may be vulnerable when a gesture consists of a combination of other similar gestures. For example, for the digit '9', an error may occur because it can be written using a combination of '0' and '1'. The confusion can be avoided if the recognition algorithm also takes into account the overall shape of the gesture trajectory.

Recently, studies on deep Convolution Neural Network (CNN) based gesture recognition have been reported. Neverova et al. proposed a multi-scale and multi-modal deep learning architecture to spot and recognize continuous gestures based on CNN [20]. Wang proposed a method of modeling spatio-temporal information occurring in a short interval by learning the motion flow map of the depth image with CNN [30]. For continuous gesture recognition purposes, 3D CNN, which is capable of learning both the spatial and the temporal aspects of the data has been applied [4]. In most of CNN framework, CNN focused on for modeling of spatial feature (e.g. hand shape) and temporal feature in short interval (e.g. motion flow). Although hand trajectories may serve as a powerful source of information in gesture recognition, to our knowledge there have been no explicit attempts to model hand trajectories using CNN.

Therefore, in this study, we propose a hand trajectory shape representation method which captures more information than conventional feature state transition in the temporal domain for continuous hand gesture recognition systems. By adding the spatial information generated from shape representation, the recognition performance of a simultaneous segmentation and recognition system can be enhanced. Specifically, we propose a segmentation method of trajectory that generates candidate trajectory segments of varying duration. In this process, shape representation provides the starting and end points of gesture strokes and curves. By improving the segmentation process, it is possible to generate segment candidates that contain fundamental characteristics of the gesture trajectories, such as temporal location and tendency of angular change. Secondly, we propose a simple but effective recognition method for gesture trajectory candidates by adapting the shape representations of total-trajectory-shape modeling to conventional temporal-feature-state transition modeling. The total-trajectory-shape modeling of gesture trajectories is similar to methods for the recognition of handwritten characters in many aspects. Since CNN have demonstrated excellent performance in handwritten digit recognition [8], we built CNN architecture suitable for total-trajectory-shape recognition and fused it with CRF-based temporal-feature modeling. The first step provides candidate gesture segments, while the second step considers the possible candidates and confirms the correct candidates for recognition. The second step thus in effect simultaneously performs final segmentation and recognition.

The remainder of this paper is organized as follows. Section 2 presents the methodology of the main proposal and its application to continuous gesture streams. The proposed methods were validated experimentally on a set of free-space drawn numeric character gestures in Section 3. Future work and concluding remarks are given in Sections 4 and 5, respectively.

## 2. Shape information based continuous gesture recognition

### 2.1. Summary of the proposed method

The overall process of the proposed system is illustrated in Fig. 1. The process is divided into three major parts: segment candidate generation (SCG), gesture classification based on trajectory shape information, and sub-gesture handling (SGH). With continuous gesture recognition, a single data stream may contain consecutive gestures, which is handled in the SCG part with the formation of a set of candidate trajectory segments. Then, in the gesture classification part, the candidates of segmented trajectories are validated by scoring to determine the best gesture class or non-gesture segments. Those candidate segments that are below the configured threshold are recognized as non-gestures, while the gesture class with the highest score is processed further. Finally, in the SGH part, gesture overlapping issue is resolved by determining whether the best gesture class is a single gesture or a sequence of multiple gestures.

### 2.2. Segment candidate generation (SCG)

As mentioned previously, the main challenge due to the issue of location ambiguity is to accurately determine intended gesture

Please cite this article as: C. Yang et al., Continuous hand gesture recognition based on trajectory shape information, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.05.016