Pattern Recognition 47 (2014) 1586–1601

Contents lists available at ScienceDirect

Pattern Recognition

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

Rule-based trajectory segmentation for modeling hand motion trajectory

Jounghoon Beh [a], David Han [b], Hanseok Ko [a,c,*]

[a] University of Maryland, College Park, MD, United States
[b] Office of Naval Research, VA, United States
[c] Korea University, School of Electrical Engineering, Anam-dong, Sungbuk-ku, Seoul 136-710, Republic of Korea

ARTICLE INFO

Article history:
Received 17 October 2012
Received in revised form
11 August 2013
Accepted 8 November 2013
Available online 25 November 2013

Keywords:
Trajectory segmentation
Hand gesture recognition
Hidden Markov model
HMM initialization

ABSTRACT

In this paper, we propose a simple but effective method of modeling hand gestures based on the angles and angular change rates of the hand trajectories. Each hand motion trajectory is composed of a unique series of straight and curved segments. In our Hidden Markov Model (HMM) implementation, these trajectories are modeled as a connected series of states analogous to the series of phonemes in speech recognition. The novelty of the work presented herein is that it provides an automated process of segmenting gesture trajectories based on a simple set of threshold values in the angular change measure. In order to represent the angular distribution of each separated state, the von Mises distribution is used. A likelihood based state segmentation was implemented in addition to the threshold based method to ensure that the gesture sets are segmented consistently. The proposed method can separate each angular state of the training data at the initialization step, thus providing a solution to mitigate the ambiguities on initializing the HMM. The effectiveness of the proposed method was demonstrated by the higher recognition rates in the experiments compared to the conventional methods.

1. Introduction

Trajectories are widely used as distinct features for recognizing not only hand gestures [1–4], but also for action recognition [5] or hand-written character recognition [6–8]. As surveyed by Rao, Yilmaz, and Shah [5], the human perception of human motion has also been shown to be based on trajectory derived information, such as the changes in the speed or direction of trajectories. In studies of hand gesture recognition, the trajectory data of moving points, such as the centroid of hands, was used for motion representation.

A gesture trajectory can be decomposed into several strokes and curves that maintain a consistent angular tendency. To accurately extract the spatio-temporal characteristics of hand movement, there have been a number of approaches involving sequential data modeling methods such as the Hidden Markov Model (HMM) [2–4,9–12], conditional random fields [13,14], hidden conditional random fields [15], input–output HMM [16], latent-dynamic conditional random fields [17], and more recently the dynamic Bayesian network [1]. Among the algorithms, the HMM has been the most widely and successfully employed in the gesture recognition task. This is due to the well-constructed theories that have been developed in various fields of research and also to its expandability to continuous symbol recognition with the language models [18] adopted from speech recognition fields. The HMM is developed based on the maximum likelihood principle, so its initialization is crucial for preventing the estimate of the HMM parameters from being caught at a local maximum. This initialization includes setting the number of states and initializing the values of the parameters, such as the mean and variance of each state in the HMM. After its initialization, the stroke or curve of the hand trajectory is modeled as a state with statistical parameters such as the mean, mixture weight, and variance that are optimized by the Baum–Welch re-estimation formulae [19] with training data.

The number of states is an important parameter, because an excessive number of states can result in an over-fitting problem if the number of training samples is insufficient compared to the number of model parameters. On the other hand, if there are an insufficient number of states, the HMM's ability to discriminate correctly is in turn reduced. A common way of determining the optimal number of states is by lengthy trial and error or by an expert's empirical decision. For example, Just and Marcel [3] determined an optimal number of states of HMM by evaluating performance of HMM in terms of the recognition rate for every step in which they increased the number of states from 2 to 20.

There have been statistical approaches to reduce this lengthy trial-and-error effort involved, particularly in the applications of

* Corresponding author. Tel.: +822 329 032 39; fax: +822 329 124 50.
E-mail addresses: hsko@korea.ac.kr, hsko@umiacs.umd.edu, hanseok.ko.34@gmail.com (H. Ko).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.11.010

HMM to speech recognition. These approaches tried to find the optimal topology of the HMM by splitting the HMM states in successive steps. The Successive State Split (SSS) algorithm, proposed by Takami and Sagayama [20], sets every state to be modeled with mixture of two Gaussian distributions, and selects a split candidate between them based on a maximum likelihood criterion. However, in many cases, it is not plausible to assume that the mixture of only two Gaussians is sufficient to model the distribution of the state. Ostendorf and Singer [21] proposed the Maximum Likelihood SSS (ML-SSS). It reduces the computational load by iteratively evaluating the split candidates. In Li and Biswas's work [22], they split the state with the largest observation density variance and merge the two states with the closest mean. These split candidates are then optimized with EM on the entire HMM. However, their algorithm fails to detect the underlying Markov states with overlapping densities, as discussed in [23]. Even though these statistical approaches can find the optimal HMM topology in terms of the maximum likelihood criteria, their use of stopping criterion such as the Bayesian information criterion [24] has not been fully justified in the HMM case.

There has been some work regarding the stopping criterion of the splitting. Biem [25] introduced the discriminative information criterion for the model selection. For each class, the HMM topology is selected so that it maximizes the difference between the log-likelihood of that model against the average log-likelihood of all the other models. Chien and Furui [26] proposed the predictive information criterion and it was used to build their tree-structured HMM topology.

The methods mentioned so far for the optimal HMM structure were proposed in the context of speech recognition or other non-gesture related areas. In the hand gesture recognition task, Siddiqi et al. [23] proposed a state splitting algorithm based on the expectation–maximization algorithm and demonstrated its use with the Australian Sign Language recognition task. Ulas and Yildiz [27] adopted a dynamic programming method to develop an optimal structure for the HMM. They postulated that the procedure of finding the optimal structure of the HMM was the same as the search in an HMM structure space, so they gradually incremented the number of states, measured the likelihood of the HMM and tried to find the optimal path to reach the optimal structure of the HMM.

In the field of hand written word recognition, Gunter and Bunke [8] [6] proposed a simple and efficient heuristic to determine the optimal number of states of the HMM by an iterative refinement with the validation data set considering the recognizer performance. Their approach can reduce the time required to find the optimal number of states, but can be seen as a type of trial-and-error procedure which requires the Baum–Welch re-estimation and the likelihood evaluation. Several works based on the observation length have been introduced. Zimmerman and Bunke [28] proposed a quintile based approach motivated by the Bakis method [29] to intuitively set the number of states in the HMM. Geiger et al. [29] proposed a data length based method. These methods showed better performance than the fixed number approaches or Bakis approach, but the data length has to be normalized beforehand.

Most of the methods mentioned so far are based on the use of a statistical approach for the optimal construction of the HMM and, therefore, they require a large set of training/validation data as well as computational complexity for assured convergence. Instead of relying on these expensive statistical approaches, we propose a rule based segmentation technique to simplify the HMM construction by first accurately estimating each trajectory state present in the gesture movements. The aim of this work is to develop a more efficient method of segmenting the trajectory according to the stationary tendency of the trajectory angle, in the case of targeted to hand motion trajectory recognition, in order to accurately set the number of states and initialize them.

Many researchers in this field have considered the use of an automated process of hand trajectory segmentation as a sub-algorithm in the recognition process. Among these methods, the segmentation of the trajectories by considering the changes in both the direction and speed of the hand motions is one of the most frequently used. Sagawa and Takeuchi [30] proposed a segmentation method using the local minima of the hand velocity and the local maxima of the angular change. In this method, the temporal boundaries specifying the sign words in a sentence are obtained by the segmentation algorithm and are fed into a sentence recognizer for the Japanese Sign Language. Wang et al. [31] employed a similar method to Sagawa and Takeuchi's approach [30] to segment the trajectory. They split the entire gesture into a so-called gesture atom as a fundamental unit of a gesture symbol. Rao et al. [5] proposed a method which segments the trajectories by examining their spatio-temporal curvature. With these boundary points, they extract dynamic instances to specify the given human action consisting of the time-frame, hand location, and the direction of hand rotation (clockwise or anti-clockwise). Gibet et al. [32] used the product of the curvature and speed of the hand movement to detect the trajectory boundary points. Kong and Ranganath [33] also segmented the trajectories in terms of their angle and speed. To minimize false segmentation due to variations of the hand movement speed, the detected boundary points were further processed by some heuristic rules.

As some of the authors of the angle-and-speed based segmentation method observed, false segmentation points often occur, due to speed variations, and it has also been noted that accurate measurements of the curvatures are dependent on the speed. To avoid these segmentation inaccuracies associated with measuring the motion speed, we propose a method based on angular features composed of geometrical figures such as lines and curves. In the experiments conducted, it was shown that these features are sufficient in terms of correctly dividing the points in a trajectory associated with significant transitions. With this novel method, the gesture motions are automatically segmented and separate motion subsets are accurately captured for the purpose of designing an optimal HMM suitable for gesture recognition.

In this work, the proposed method can be effectively employed for initializing the HMM for hand gesture recognition. Specifically, the proposed method is applied to the training phase to determine the number of states and its initial estimates of the probability distribution parameters. Our proposed algorithm consists of two-steps, as depicted in Fig. 1. First, a novel method of motion segmentation termed the Figure-based Trajectory Segmentation (FTS) algorithm, enhanced from our previous work [34], is proposed to automatically determine the number of states based on a simple set of rules in terms of the trajectory angles. Since the hand trajectory has several turning points and consists of straight lines and curves, if each figure can be properly divided, it can help to efficiently model the HMM. However, the number of segments and their location might not be consistent throughout the training data, due to the variant realization of the hand trajectory present in the training data. Hence, we augment a method of realigning the segmented points in a trajectory. In this way, our approach provides a robust HMM initialization method for improving the performance of hand gesture recognition.

For a gesture recognition system, the trajectory feature of the hand movement can be captured in numerous ways: data gloves [18] [35], 3D sensors [36,37], and multiple cameras [38,39]. In this work, we focus on the recognition of the hand trajectory with the assumption that a suitable hand detection/tracking algorithm has already been applied to extract the trajectory. Note, therefore, that we do not attempt to deal with the hand detection and tracking task.

The remainder of this paper is organized as follows. In Section 2, the basic concept of the proposed segmentation method is described