Signal Processing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Signal Processing

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

Online gesture-based interaction with visual oriental characters based on manifold learning

Yi Wang [a], Juncheng Liu [b], Xin Fan [a,*], Xiangjian He [c], Qi Jia [a], Renjie Gao [a]

[a] School of Software, Dalian University of Technology, Dalian 116620, China
[b] School of Electronics Engineering and Computer Science, Peking University, Peking 100000, China
[c] School of Computing and Communications, University of Technology, Sydney, Australia

A R T I C L E I N F O

Article history:
Received 28 April 2014
Received in revised form
23 August 2014
Accepted 29 August 2014

Keywords:
Human–computer interface
Visual oriental characters
Feature descriptors
Graph-optimized locality preserving projections

A B S T R A C T

Online gesture-based interaction with characters has become a more natural and informative human–computer interface with the popularity of new interactive devices (e.g., Kinect and Leap Motion). In this paper, a new feature descriptor named Segmented Directed-edge Vector (*SDV*) is proposed. This simple and yet quite effective descriptor is able to capture the characteristics of visual oriental characters. Moreover, we explicitly build the mappings from *SDVs* to features in a subspace by a modified Locality Preserving Projections (LPP) method with stroke class constraints. These mappings can yield meaningful subspace structures for larger character sets. Extensive experiments on the online interactive system demonstrate the robustness of our method to various issues in gesture-based character's input, such as unnatural breaks, overlapped or distorted radicals, and unconscious or quivering trajectories. Our system can still achieve accurate recognition when accumulative errors occur with complex characters.

1. Introduction

One of the most attractive means of natural human computer interaction (HCI) is gesture-based character input and recognition (GCIR) that provides more interactive controls [1]. GCIR has great potentials in real applications, such as disable assistance, interactive computer games, smart appliances and writing trainings. In the field of computer vision (CV), GCIR can be regarded as the merge of character recognition and gesture recognition. Researchers have developed numerous approaches in both areas [1–5], but their feature extractors and classifier learning algorithms may be inapplicable to GCIR. On one hand, various 'static' features, such as gradient features, Gabor features and SIFT, are widely used for handwritten character recognition [6–8]. Nevertheless, radicals or strokes of a character are likely to be overlapped or distorted in the context of GCIR that is different from the traditional way of writing, does not have a writing plane (paper or touch screen) and does not have any demarcation for writing. On the other hand, classical classifiers including Hidden Markov Models (HMM) [9] and Conditional Random Field (CRF) [10] provide an elegant framework to label gestures. Unfortunately, these methods can hardly render the real-time performance for larger label sets, when there are, for example, thousands of or more gestures (characters). GCIR also demands an incremental learning ability that embraces new characters and writing styles in an online fashion. Furthermore, learning strategies in GCIR have to be robust to outliers brought by unconscious quivering trajectories and accumulative errors.

In this paper, we put forwards a novel online interactive character's input and recognition method by gestures based on manifold learning. We utilize the directional and ordered information of gestures in GCIR to design a new descriptor that has a good discriminative ability in recognition.

* Corresponding author. Tel.: +86 411 87571638.
E-mail address: xin.fan@ieee.org (X. Fan).

http://dx.doi.org/10.1016/j.sigpro.2014.08.042
0165-1684/

We study subspace learning techniques and learn explicit mappings to recover intrinsic configurations for numerous characters, and obtain real-time and high recognition rate. The main contributions of this paper are given below:

(1) We propose a new stroke-level feature representation for dynamic gestures, named Segmented Directed-edge Vector (*SDV*). *SDV* is invariant to scale variations, while combining both geometrical and dynamical information of trajectories. *SDVs* have a low repetition rate on a large character set, and have a reasonable tolerance of writing fault.
(2) We use *SDVs* and a modified Locality Preserving Projections (LPP) [11,12] to obtain robust and efficient recognition results on large character sets. Stroke class-specific constraints are incorporated in order to generate more meaningful subspace structures and explicit mapping functions for classification. We are able to achieve real-time recognition by applying a parallel searching technique.
(3) We implement the proposed practical gestures-based character input and recognition (GCIR) system using depth images from monocular videos captured by a Kinect sensor and the features obtained in (2) above. This system can tackle unnatural breaks, overlapped or distorted radicals, and unconscious or quivering trajectories in real applications. Moreover, we build a semantic associative database that can accelerate the speed of the system further.

2. Related work

In recent years, studies in the computer vision area on motion recognition and shape reconstruction have witnessed a growing interest in subspace analysis and manifold learning techniques [13–15]. Given a set of high-dimensional data points, manifold learning techniques aim at discovering the geometric properties in a data space, including the Euclidean embedding, intrinsic dimensionality, connected components and homology. Manifold learning techniques can be classified into linear (MDS [16], LDA and PCA [17], etc.) and non-linear (ISOMap [18], LLE [19], LE [20], RML [21], LTSA [22], etc.) methods, which have been applied to face and gait recognition and show impressive performance.

Locality Preserving Projections (LPP) [11] attempt to find optimal linear approximations to the Eigen-functions of the Laplace Beltrami operator on a manifold. This technique seeks to preserve the intrinsic geometry of data and the local structures. In contrast to other manifold learning algorithms, LPP possess a remarkable advantage that can generate an explicit mapping. Furthermore, compared with HMM or CRF methods, LPP are defined everywhere of training data and may be simply applied to any new data. Many improvements on LPP have emerged in the previous few years. Yen et al. [23] proposed an orthogonal neighborhood preserving discriminant analysis (ONPDA) method, which effectively combines the characteristics of LDA and LPP. Wong and Zhao [24] proposed two feature extraction algorithms derived from LPP, i.e., the supervised optimal locality preserving projection (SOLPP) algorithm and the normalized Laplacian-based supervised optimal locality preserving projection (NL-SOLPP). These algorithms use both local information and class information to model the similarity of data.

The construction of a neighbor weight graph is the key to subspace learning algorithms [11–15]. Recent studies show the influence of the graph construction on clustering measures and resultant subspace representation. Traditional construction methods using k-nearest neighbors typically lead to an unbalanced graph and thus unfavorable performance. Zhang et al. [12] developed an unsupervised Graph-optimized Locality Preserving Projections (GoLPP), which incorporated graph construction into the LPP objective function, and thus obtained a joint learning framework for graph construction and projection optimization. GoLPP produce a changeable graph instead of a fixed one in LPP. The graph is gradually updated in an iterative process, and naturally takes transformed data. However, the adjacent graph of GoLPP is initialized by the traditional k-nearest neighbor method without class information. Therefore, the initialization may take some heterogeneous samples for optimization. Yu et al. proposed a hypergraph-base manifold learning in [25]. By varying the neighborhood size, they generated a set of hyperedges for each image and its visual neighbors. The joint learning of the image labels and hyperedge weights automatically modulates the effect of hyperedges. In [26], they adopted sparse representation to select a few neighbors of each data point that span a low-dimensional affine subspace passing near that point. After that, the whole alignment strategy is utilized to build the manifold.

3. Framework of the proposed approach

In this paper, we focus on developing a gesture based and real-time recognition method with interactive input on large visual oriental character sets. We name the trajectory of a control hand from the beginning to the end in the process of writing a character, including the transition edges between strokes, as a Visual Oriental Character (VOC). Oriental characters refer to stroke and structure based characters, such as the characters in Chinese, Japanese and Korean. Examples for structured oriental characters are shown in Fig. 1. Besides, English letters and Arabic numerals can also be written in this style using gestures.

Our approach consists of four steps, i.e., control hand recognition and tracking, directed-edges quantified and optimization, feature extraction and selection, and large dataset training and classification, as shown in Fig. 2. In the off-line pre-processing stage, *SDVs* are computed for VOCs in training sets. Then, a subspace learning algorithm is adopted to learn explicit mappings of features from high-dimensional feature spaces to features in low-dimensional embedding spaces. Moreover, a semantic associative database is built to accelerate the speed of input. In the on-line processing stage, we segment the control hand from depth images obtained from monocular videos captured by a Kinect sensor. And a stroke tracking

Please cite this article as: Y. Wang, et al., Online gesture-based interaction with visual oriental characters based on manifold learning, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.08.042