Online unconstrained handwritten Tibetan character recognition using statistical recognition method

Permalink
https://escholarship.org/uc/item/3dk73778

Himalayan Linguistics, 15(1)

Ma, Long-long
Wu, Jiang

2016

10.5070/H915130066

CC BY-NC-ND 4.0

Peer reviewed
Himalayan Linguistics

Online unconstrained handwritten Tibetan character recognition using statistical recognition

Long-Long Ma  Jian Wu
Chinese Academy of Sciences

ABSTRACT

This paper describes a recognition system for online handwritten Tibetan characters using advanced techniques in character recognition. To eliminate noise points of handwriting trajectories, we introduce a de-noising approach by using dilation, erosion, and thinning operators of mathematical morphology. Selecting appropriate structuring elements, we can clear up large amounts of noise in the glyphs of the character. To enhance the recognition performance, we adopt a three-stage classification strategy, where the top rank output classes by the baseline classifier are re-classified by a similar character discrimination classifier. Experiments have been carried out on two databases MRG-OHTC and IIP-OHTC. Test results show the recognition algorithm employed is effective and can be applied to pen-based mobile devices.

KEYWORDS

online handwritten Tibetan character recognition, de-noising, pre-processing, three-stage classification
Online unconstrained handwritten Tibetan character recognition using statistical recognition

Long-Long Ma                    Jian Wu
Chinese Academy of Sciences

1 Introduction

Tibetan is a language with a long history of over 1,300 years. The Tibetan language is still used by more than six million people in China, especially in the Tibet Autonomous Region (Xizang), Yunnan and Qinghai provinces. Research on Tibetan characters, which will facilitate the engagement of Tibetans with modern technologies and enable the digitization of Tibetan documents, is very important both from theoretical and practical perspectives.

Due to the increase of new pen input devices and pen applications, online handwritten character recognition is gaining increasing interest. However, compared to the existing research work on CJK (Chinese, Japanese and Korean) and Arabic, online handwritten Tibetan character recognition (OHTCR) is a relatively unexplored field.


This paper describes an online recognition system for handwritten Tibetan characters and reports our experimental results using an approach based on de-noising and a three-stage classification strategy. As for all handwritten recognition problems, handwritten Tibetan character recognition is difficult due to the wide variability of writing styles and the confusion between similar characters. The methods of online character recognition can be roughly grouped into two categories: statistical and structural (see Liu (2004)). Whereas structural matching is more relevant to human learning and perception, statistical methods are more computationally efficient. Taking advantage of learning from samples, statistical methods can give higher recognition accuracies.

We adopt a statistical classification scheme, wherein recognition accuracy depends on the techniques of pre-processing, feature extraction, and classifier design. We use a de-noising approach to eliminate noise points of trajectories (see Sun (2009)). Equidistance re-sampling, smoothing and
nonlinear shape normalization are used in a pre-processing step. The local stroke direction of a character pattern is decomposed into direction maps, which are blurred and sub-sampled to obtain feature values (see Hamanaka (1993) and Zhou (2009)).

The confusion between similar characters is one of the main reasons for lower recognition accuracy. The recently proposed LDA (linear discriminant analysis)-based compound distance (see Gao (2008)) and the critical region analysis based pair discrimination (see Leung (2010)) further improve recognition accuracy. A logistic regression (LR) classifier is used to discriminate confusing characters and costs only small storage for extra parameters, compared to those methods of Gao (2008) and Leung (2010). To discriminate between confusing characters, we use a three-stage classification strategy, similar to the strategy by Zhou (2010) for Japanese. Firstly, candidate classes are selected with a coarse classifier according to the Euclidean distance (ED) to class means. Secondly, fine classification with the modified quadratic discriminant function (MQDF) (see Kimura (1987)) re-orders the candidate classes. Thirdly, the top ranked candidate classes are re-classified by a similar character discrimination classifier. Our experiments on two databases, MRG-OHTC and IIP-OHTC, demonstrate that similar character discrimination classifier can improve the recognition accuracy by about 3%.

The rest of this paper is organized as follows. Section 2 describes the structural characteristic of Tibetan characters. Section 3 gives an overview of the recognition system. Section 4 describes a de-noising approach during pre-processing and section 5 introduces a three-stage classification strategy. Section 6 reports our experimental results and section 7 provides concluding remarks.

2 Structural characteristic of Tibetan characters

The Tibetan script distinguishes four vowels and 30 consonants, which may be called its basic elements. There are two kinds of characters used in handwritten Tibetan characters, that is, single characters (SC) and combined characters (CC).

Syllables are basic spelling units (see Ding (2007)), whose structure is shown in Fig.1. Each syllable may consist of up to four characters (those parts surrounded by red dash line boundary boxes in Fig.1). The combined characters, which consist of at least an essential consonant (EC, in Tibetan called ming gzhi) and may include a top vowel (TV), a consonant above the EC (CaEc, mgo can), a consonant below the EC (CbEc, ’dogs can), and a bottom vowel (BV). Some consonants (in total 20) can serve as a character located to the left of the CC (CbCC, sngon’ jug), located or the immediate right of the CC (1-CaCC, rjes’ jug), or located two positions to the right of the CC (2-CaCC, yang’ jug). Fig.2 gives an example of a four character syllable. In this paper we only consider isolated Tibetan character (SC and CC) recognition.

Figure 1. The syllable structure
3 System overview

The OHTCR system is shown diagrammatically in Fig. 3. The input pattern trajectory is composed of the coordinates of sampled pen-down points. To eliminate noise points and remove certain variations among character samples of the same class, we introduce a de-noising approach at the pre-processing stage. For feature extraction, we use a direction feature extraction method (see Hamanaka (1993)) where the stroke direction is the one in the original pattern, not in the normalized pattern. With the three-stage classification process, we introduce similar character discrimination to reduce the recognition error caused by confusion between similar characters.

4 De-noising pre-processing

Pre-processing is used to regulate the pattern shape for reducing the class internal shape variation. Nonlinear shape normalization (NSN) is used to normalize shape variability (see Bai (2006)). However, the proportion of noise points in character point trajectories is magnified after NSN. Fig. 4 shows two examples. It is important to eliminate these noises in order to permit these characters to be recognized accurately.
To eliminate these noise points, we introduce a de-noising method before NSN. The pre-processing process is illustrated in Fig. 5, where the de-noising step employs the operators (dilation, erosion and thinning) of mathematical morphology. Linear size transformation is to ensure that the character samples of the same class have approximately the same size. Re-sampling is used to reduce distance variation between two adjacent online points. We use Gaussian smoothing (Gaussian blur) to reduce stroke variation in a small local region. We take the image composed of trajectory points as the binary image.

4.1 Dilation and erosion

Assuming that $A$ is a region in a binary image $I$, and that $B$ is a structure element, the dilation of $A$ by $B$ (written $A \oplus B$) is defined as

$$A \oplus B = \left\{ p \left| B_{\hat{}} \cap A \neq 0 \right. \right\}$$

(1)

where $\hat{B}$ is the symmetric of $B$ and $B_{\hat{}}$ is the translation of $B$ by the vector $\hat{p}$. Under the same assumptions, the erosion of $A$ by $B$ (written $A \ominus B$) is defined as

$$A \ominus B = \left\{ p \left| B_{\hat{}} \subseteq A \right. \right\}$$

(2)

Using the dilation operation, some useless or re-written strokes can be connected to a component. Unlike offline recognition, online handwriting records stroke direction and point time sequences. According to the point time information, we dilate the binary image. For the erosion operator, we use the same size structuring element (a 3×3 square, with the origin at its center) that is
used for the dilation operator. Fig. 6(a) and 6(b) respectively give the dilation and erosion results of an original pattern.

4.2 Thinning

After dilation and erosion operators, we get the images composed of the strokes with different pixel widths, which is shown in Fig 6(b). In order to regulate the transformed image, we use the thinning algorithm (see Zhang (1984)) to extract a stroke skeleton. Finally the binary image consists of strokes with one pixel width. Compared to the original pattern, we can see the noise points are removed from Fig. 6(c).

Fig. 7 gives the transformed image after pre-processing with and without the de-noising method. We can see the image (c) in shape is more similar to the corresponding intended shape.
5 Three-stage classification

After pre-processing and feature extraction of the input pattern, the feature dimensionality is reduced by LDA. The coarse classifier gives some candidate classes according to the ED from the reduced vector, and the fine classifier (MQDF as the baseline classifier) reorders the candidate classes.

The similar character sets are built on the training dataset by 5-fold cross validation, i.e. rotationally using 4/5 for training the baseline classifier and the remaining 1/5 for validation. We use the selection criterion proposed by Gao (2008) to get the similar character sets.

Assuming the baseline classifier outputs a ranked candidate list. In our experiment, we select the top 5 outputs. If any two of the candidate list belong to one of the similar character sets, we use two-class LDA (see Gao (2008)) to discriminate between them. In LDA, the projection axis \( w \) for discriminating two classes is estimated to maximize the Fisher criterion. The optimal discriminant vector is represented as

\[
\mathbf{w} = \mathbf{S}_i^{-1} \left( \mathbf{u}_i - \mathbf{u}_j \right) = \mathbf{S}_i^{-1} \left( \mathbf{u}_i - \mathbf{u}_j \right)
\]

(3)

where \( \mu_i \) and \( \mu_j \) are the means of two classes. \( \Sigma_i \) is the average covariance matrix of two classes and can be rewritten as

\[
\mathbf{w} = \mathbf{S}_i^{-1} \left( \mathbf{u}_i - \mathbf{u}_j \right) = \mathbf{S}_i^{-1} \left( \mathbf{u}_i - \mathbf{u}_j \right) = \sum_{m=1}^{d} \frac{1}{\lambda_n} \mathbf{y}_n^T \mathbf{y}_n \sum_{m=1}^{d} \mathbf{y}_n^T \mathbf{y}_n \mathbf{y}_n ^T
\]

(4)

Finally the similar character discriminant function is formulated as

\[
f(x) = f_{LDA}(x) = \mathbf{w}^T \mathbf{x}
\]

(6)

For the top \( N \) output classes from the baseline classifier, we can get at most \( N \times (N-1)/2 \) classification results after similar character discrimination. The final decision is determined by majority voting.

6 Experimental results

We evaluated the recognition performance on two databases of online handwritten Tibetan characters: MRG-OHTC and IIP-OHTC. The MRG-OHTC, collected by our research group, contains handwritten Tibetan samples of 910 character classes, 130 samples for each character class. We choose the first 105 samples from each class for training, and the remaining 25 samples from each class for testing. The IIP-OHTC database, collected by the Northwest University for Nationalities, contains 562 characters, 150 samples for each character class. We choose 120 samples per class for training and the remaining 30 samples per class for testing.

For each character pattern, we extract a 512-dimensional directional features (see Hamanaka (1993)). The 512-dimensional feature vector is projected onto a 140-dimensional subspace learned by global LDA. The 140-dimensional projected vector is then fed to the MQDF classifier (fine classification), with 40 principal eigenvectors for each class. Table I lists the recognition accuracy for the two databases using the baseline (MQDF) classifier.
From Table I we can see that the test accuracy is lower on the two databases, and there is a large accuracy difference between top1 and top2, and between top2 and top5. The inaccuracy is mainly attributable to the confusion between similar characters. We use a two-class LDA to further identify similar characters. The recognition accuracy improves about 3%. Obviously, it is very challenging to present new algorithms for higher accuracy.

<table>
<thead>
<tr>
<th></th>
<th>MRG-OHTC</th>
<th>IIP-OHTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top1</td>
<td>81.70%</td>
<td>77.01%</td>
</tr>
<tr>
<td>Top2</td>
<td>90.96%</td>
<td>88.99%</td>
</tr>
<tr>
<td>Top5</td>
<td>96.04%</td>
<td>95.79%</td>
</tr>
</tbody>
</table>

Table 1. Test Accuracy

Fig.6 shows the samples misrecognized by MQDF, but corrected by two-class LDA discrimination classifier. Fig.8 gives some examples with 5 candidate outputs, where the correct results are labeled in red. We can see these five candidates are very similar in shape. The misrecognized results cannot be corrected using the two-class LDA.

Figure 8. Examples of misrecognized characters corrected by similar character discrimination

Figure 9. Examples of the top 5 candidate outputs

Though the correct rate on two databases is lower, the accumulated recognition rate of the top 10 is higher than 97%. We apply the recognition algorithms to pen-based applications such as mobile phones. Fig.10 shows the interface of our recognition system, where the left sub-window displays the trajectories of handwriting, and the right sub-window gives the top ten recognition results. When the correct recognition result is selected, the character class is surrounded by red bounding box. The bottom sub-window gives the character string with correct outputs.
7 Conclusion

In this paper, we describe a recognition system for online handwritten Tibetan characters. At the pre-processing step, the de-noising method is used to eliminate the noise points of character trajectories. A three-stage classification strategy reduces the error from the confusion between similar characters. The experiments on MRG-OHTC and IIP-OHTC databases demonstrated the recognition algorithm can be applied as a real recognition system. To further improve the recognition system, we are considering more recognition algorithms and a strategy of combining multiple classifiers.

Acknowledgements

This work is supported by the CAS Action Plan for the Development of Western China (No.KGCX2-YW-512) and National Science & Technology Major Project (No.2010ZX01036-001-002, 2010ZX01037-001-002). The authors would thank Professor Weilan Wang from Northwest University for Nationalities, for providing the IIP-OHTC database.

References

Bai, Zhen-long; and Huo, Qiang. 2006. “A study of nonlinear shape normalization for online handwritten Chinese character recognition: dot density vs. line density equalization”. In: Proceedings of the 18th international conference on pattern recognition (ICPR), pp. 921-924.
Ding, Xiaoqing; and Wang, Hua. 2007 “Multi-font printed Tibetan OCR”. In: Chaudhuri, Bidyut B. (Ed.) Advance in pattern recognition, 73-98. London: Springer.
Ma and Wu: Online unconstrained handwritten Tibetan character recognition


Ngodrup; and Zhao, Dongcai. 2010. “Study on printed Tibetan character recognition”. In: Proceedings of Artificial Intelligence and Computational Intelligence (AICI), pp. 280-285.


Long-Long Ma
longlong@iscas.ac.cn