2009 10th International Conference on Document Analysis and Recognition

Affine Distortion Compensation for an Isolated Online Handwritten Chinese Character Using Combined Orientation Estimation and HMM-based Minimax Classification

Tingting He[1] and Qiang Huo[2]
[1]Department of Computer Science, The University of Hong Kong, Hong Kong, China
[2]Microsoft Research Asia, Beijing, China
(E-mails: tthe@cs.hku.hk, qianghuo@microsoft.com)

Abstract

This paper presents a new approach to compensating affine distortion of an isolated online handwritten Chinese character. The input sample is first analyzed by using a character-structure-guided orientation estimation approach. If necessary, the orientation hypotheses are refined based on confidence evaluation of two pre-classifiers. Depending on the number of possible orientations, an HMM-based minimax classification approach is then used to estimate an affine transformation against either the original sample or the compensated sample with the previously identified orientation. The final compensated sample can be derived accordingly using the estimated affine transformation. The effectiveness of the proposed approach is demonstrated by recognition experiments using distorted samples generated artificially from the popular Nakayosi and Kuchibue Japanese character databases.

1 Introduction

After many years of research, it is still an unsolved research problem to construct an online handwritten Chinese character recognition (HCCR) system which is robust against the possible global affine distortions of an *isolated* input handwriting sample. In [7], we proposed to use a minimax classification rule to recognize an unknown Chinese character with a possible affine distortion, where each character is modeled by several continuous-density hidden Markov models (CDHMMs) with "coordinates difference" extracted from each point of the fine trajectory of the handwriting sample as a feature vector, and a global affine transformation for all the mean vectors of each CDHMM is used to serve as a structural distortion model. Although very promising results have been achieved and demonstrated in [7], doing minimax classification for large vocabulary HCCR is computationally too expensive. Furthermore, a state-of-the-art online HCCR system typically consists of at least two types of classifiers, one is based on techniques such as HMM approach which can make good use of the temporal information in a handwriting sample, and another is based on traditional classifiers such as MQDF-based classifiers [8] or multiple-prototype (MP) based classifiers (e.g., [4]) constructed with features (e.g., [1, 2]) insensitive to the variability of writing orders of strokes. Apparently, it is important to make the second type of classifiers robust against affine distortions as well. It is the purpose of this paper that we propose a new approach to address this problem.

The rest of the paper is organized as follows. In Section 2, we present our new approach in detail. The experimental results are then reported in Section 3. Finally, we conclude this paper in Section 4.

2 Our Approach

In this paper, we propose to compensate the affine distortion of an isolated online handwritten Chinese character using combined orientation estimation and HMM-based minimax classification. Each unknown and possibly distorted handwriting sample is first analyzed by using a character-structure-guided orientation estimation approach [5]. The orientation hypotheses are further refined based on confidence evaluation of two pre-classifiers. Depending on the number of possible orientations, the set of aforementioned CDHMMs and the minimax classification approach are then used to estimate an affine transformation against either the original sample or the compensated sample with the previously identified orientation. With the estimated affine transformation, the final compensated sample can be derived accordingly. To speed up the estimation process, the minimax classification is conducted on a short-list of candidates hypothesized by an efficient single-prototype (SP) based classifier which is robust against rotation distortions with the help of orientation compensation [5]. The overall flow chart of our proposed approach is illustrated in Fig. 1.

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.87

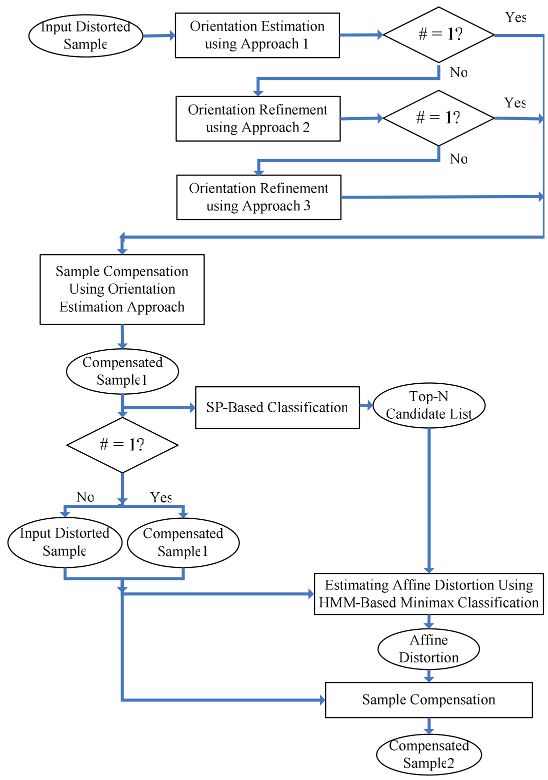

Figure 1. Overall flow chart of the approach compensating affine distortions using combined orientation estimation and HMM-based minimax classification.

In the following subsections, we elaborate on how to generate the "Top-N" candidate list, how to use CDHMM-based minimax classification approach to estimate the possible affine distortion, and how to estimate and refine orientations from the input handwriting sample, respectively.

2.1 Generation of Top-N Candidate List

Given a possibly distorted handwritten character sample, after preprocessing, it can be represented as a time-ordered sequence of points, $\mathbf{P} = (P_0, P_1, \cdots, P_t, \cdots, P_T)$, where $P_t = (x_t, y_t)$ is the coordinates of the t-th point. Then, the orientation estimation approach to be discussed later is used to estimate L ($L \leq 3$) possible character orientations denoted as $\{\phi_i\}_{i=1}^{L}$. For each hypothesis ϕ_i, the input sample $\mathbf{P}$ is then compensated as $\hat{\mathbf{P}}_i$ by transforming each sample point with the following transformation matrix:

$$W_{\phi_i} = \left[\begin{array}{cc} \cos(-\phi_i) & -\sin(-\phi_i) \\ \sin(-\phi_i) & \cos(-\phi_i) \end{array} \right] .$$

Then, a 512-dimensional raw feature vector is extracted from each compensated sample $\hat{\mathbf{P}}_i$ using the approach in [2] which is an improved version of the approach proposed originally in [1]. A new 80-dimensional feature vector is then obtained via LDA (Linear Discriminant Analysis) transformation estimated from training samples. A single-prototype (SP) based classifier, where the prototype for each character class is estimated as the sample mean of the corresponding 80-dimensional training feature vectors, can be used to generate a short-list of top N ($N \geq 3$) candidates for each unknown and possibly distorted sample $\mathbf{P}$, where each compensated sample $\hat{\mathbf{P}}_i$ contributes $\frac{N}{L}$ candidates different from the ones provided by other compensated samples.

2.2 Estimating Affine Distortion Using HMM-Based Minimax Classification

Given a handwriting sample $\mathbf{P}$, it can be represented as a sequence of observation feature vectors $\mathbf{O} = (O_1, O_2, \cdots, O_T)$, where $O_t = (x_t - x_{t-1}, y_t - y_{t-1})^{Tr}$ is a simple 2-dimensional feature vector computed at point P_t. We use Gaussian-mixture CDHMM to model the whole character directly for simplicity. Let's assume that there are M character classes, C_i, $i = 1, 2, ..., M$. In order to deal with the variability of writing orders of strokes, the set of non-distorted training data for each character class C_i is clustered into several subsets. For the n-th subset with sufficient amount of training data, a left-to-right CDHMM, $\lambda_i^{(n)}$, allowing state transitions of skipping one state, is constructed by using maximum likelihood (ML) training. The number of states, $NS_i^{(n)}$, in each CDHMM $\lambda_i^{(n)}$ is determined as the number of line segments of a representative sample in the corresponding subset of training data. Consequently, each character class, C_i, will have N_i CDHMMs, $\{\lambda_i^{(n)}, n = 1, \cdots, N_i\}$.

As described in [6, 7], a *minimax classification rule* can be used to classify the unknown character sample $\mathbf{O}$ as class C_i, if

$$i = \arg\max_j \{\max_n [\max_S \log p(\mathbf{O}, S | \mathcal{T}_{\hat{W}}(\lambda_j^{(n)}))]\} \quad (1)$$

where $p(\mathbf{O}, S | \mathcal{T}_{\hat{W}}(\lambda_j^{(n)}))$ is the joint likelihood of the observation $\mathbf{O}$ and the associated hidden state sequence S given the transformed CDHMM $\mathcal{T}_{\hat{W}}(\lambda_j^{(n)})$ with the parameters $\hat{W}$ estimated as follows:

$$\hat{W} = \arg\max_W p(\mathbf{O} | \mathcal{T}_W(\lambda_j^{(n)})) \ . \quad (2)$$

In Eq. (1), the operation "$\arg\max_j$" is conducted over the "Top-N" candidate list identified as described in the previous subsection. A global transform $\mathcal{T}_W(\cdot)$ is used and applied to each mean vector of a CDHMM as $\hat{\mu} = W \cdot \mu$, where W is a 2×2 matrix, and μ is a 2-dimensional mean vector of a particular Gaussian component of the CDHMM. W can be solved in Eq. (2) by using an iterative EM algorithm [7], which is a special case of a technique known as maximum likelihood linear regression (MLLR) [9] in speech recognition field. The optimal affine transform $\hat{W}$ for the best recognition result can then be used to compensate $\mathbf{P}$ by transforming the coordinates of each point as follows:

$$(\hat{x}_t, \hat{y}_t)^{Tr} = \hat{W}^{-1} \cdot (x_t, y_t)^{Tr} \ .$$

The compensated sample can be fed into a traditional MP-based or MQDF-based classifier for recognition.

2.3 Estimating and Refining Orientations

In [5], we have proposed a character-structure-guided approach (referred to as "Approach 1" hereinafter) to estimate a maximum of 3 possible character orientations from a distorted handwriting sample. In the following, we present two approaches to further reducing the number of hypothesized orientations based on confidence evaluation of two different classifiers.

2.3.1 Reducing Hypothesized Orientations via Confidence Evaluation of an SP-based Classifier

For a given distorted sample $\mathbf{P}$, "Approach 1" is used first to estimate a set of possible orientations $\{\phi_i\}$. Then, the SP-based classifier is used, as described in section 2.1, to recognize the compensated samples $\{\hat{\mathbf{P}}_i\}$ with the hypothesized orientations and generate the "Top N" candidate list. Let $\tilde{S}_i^j$ denote the " j-th best candidate" recognition score of the i-th compensated sample, where the Euclidean distance between the input feature vector and the corresponding "prototype" is used as the recognition score. For the convenience of discussions, ϕ_i's have been sorted in ascending order of $\tilde{S}_i^1$'s.

If "Approach 1" generates two orientation hypotheses, the following confidence measure (CM) is defined:

$$CM = \begin{cases} 1 & |\tilde{S}_2^1 - \tilde{S}_1^1| > \delta \quad or \\ & |(\tilde{S}_1^2 - \tilde{S}_1^1) - (\tilde{S}_2^2 - \tilde{S}_2^1)| > \eta \\ 0 & otherwise \end{cases}$$

where δ and η are two control parameters tuned empirically on development set to achieve the best performance ($\delta = 1.8$ and $\eta = 1.2$ in our experiments). The orientation hypotheses can then be refined by using the following rule:

- If $CM = 1$, only ϕ_1 is retained; Otherwise, no change is made to the orientation hypotheses.

If "Approach 1" generates three orientation hypotheses, the following two confidence measures are defined:

$$CM_1 = \begin{cases} 1 & |\tilde{S}_2^1 - \tilde{S}_1^1| > \delta_1 \quad or \\ & |(\tilde{S}_1^2 - \tilde{S}_1^1) - (\tilde{S}_2^2 - \tilde{S}_2^1)| > \eta_1 \\ 0 & otherwise \end{cases}$$

$$CM_2 = \begin{cases} 1 & |\tilde{S}_3^1 - \tilde{S}_2^1| > \delta_2 \\ 0 & otherwise \end{cases}$$

where δ_1, η_1, and δ_2 are three control parameters tuned empirically on development set to achieve the best performance ($\delta_1 = 1.0$, $\eta_1 = 0.5$, and $\delta_2 = 1.5$ in our experiments). The orientation hypotheses can then be refined by using the following rules:

- If $CM_1 = 1$, only ϕ_1 is retained;
- If $CM_1 = 0$ and $CM_2 = 1$, ϕ_1 and ϕ_2 are retained;
- Otherwise, all three orientations will be kept for further processing.

The above approach to refining the set of orientations hypothesized by "Approach 1" will be referred to as "Approach 2" hereinafter. It is noted that if any change is made to the orientation hypotheses by "Approach 2", the set of "Top N" candidate list will also be adjusted accordingly.

2.3.2 Reducing Hypothesized Orientations via Confidence Evaluation of an HMM-based Minimax Classifier

If there still exist multiple orientation hypotheses $\{\phi_i\}$ even after using "Approach 2", they may be further refined by evaluating a confidence measure based on the HMM-based minimax classifier described in section 2.2. More specifically, for each compensated sample $\hat{\mathbf{P}}_i$ with the hypothesized orientation ϕ_i, a minimax classification can be conducted. Let $\breve{S}_i^1$ denote the "best candidate" recognition score of the i-th compensated sample as defined in Eq. (1). Again, for the convenience of discussions, ϕ_i's have been sorted in descending order of $\breve{S}_i^1$'s. A new confidence measure is defined as follows:

$$CM = \begin{cases} 1 & |\breve{S}_2^1 - \breve{S}_1^1| > \tau \\ 0 & otherwise \end{cases}$$

where τ is a control parameter tuned empirically on development set and set as 150 in our experiments. The orientation hypotheses can then be refined by using the following rule:

- If $CM = 1$, only ϕ_1 is retained; Otherwise, no change is made to the orientation hypotheses.

The above approach to refining the set of orientations after using "Approach 2" will be referred to as "Approach 3" hereinafter. Again, if any change is made to the orientation hypotheses by "Approach 3", the set of "Top N" candidate list will also be adjusted accordingly.

Table 1. Frequency distribution of estimation errors $\Delta\phi_{min}$ falling in each bin on 7 testing sets with various distortion degrees when using "Approach 2" for character orientation estimation.

	T_{-30}	T_{-20}	T_{-10}	T_0	T_{10}	T_{20}	T_{30}
Bin1	0.31	0.31	0.31	0.31	0.32	0.32	0.33
Bin2	0.35	0.34	0.35	0.33	0.33	0.33	0.33
Bin3	0.16	0.16	0.15	0.21	0.18	0.18	0.18
Bin4	0.10	0.10	0.11	0.09	0.09	0.09	0.12
Bin5	0.05	0.05	0.05	0.05	0.05	0.05	0.03
Bin6	0.02	0.02	0.01	0.01	0.01	0.02	0.01
Bin7	0.01	0.02	0.02	0.0	0.02	0.01	0.00
#	1.07	1.07	1.06	1.05	1.05	1.05	1.05

Table 2. Frequency distribution of estimation errors $\Delta\phi_{min}$ falling in each bin on 7 testing sets with various distortion degrees when using "Approach 3" for character orientation estimation.

	T_{-30}	T_{-20}	T_{-10}	T_0	T_{10}	T_{20}	T_{30}
Bin1	0.31	0.31	0.31	0.31	0.32	0.32	0.32
Bin2	0.35	0.34	0.35	0.32	0.32	0.33	0.33
Bin3	0.16	0.16	0.15	0.20	0.18	0.18	0.18
Bin4	0.09	0.10	0.11	0.08	0.09	0.08	0.11
Bin5	0.04	0.04	0.04	0.04	0.05	0.05	0.03
Bin6	0.02	0.02	0.01	0.01	0.01	0.02	0.01
Bin7	0.03	0.03	0.03	0.04	0.03	0.02	0.02
#	1.03	1.03	1.03	1.02	1.02	1.02	1.02

3 Experiments and Results

3.1 Experimental Setup

In order to demonstrate the effectiveness of our newly proposed approach, a series of comparative experiments have been designed and conducted on the task of the recognition of isolated online handwritten characters with a vocabulary of 2965 level-1 Kanji characters in JIS standard. The popular Nakayosi and Kuchibue Japanese character databases [10] are used here. The Nakayosi database consists of about 1.7 million character samples from 163 writers, and the Kuchibue database contains about 1.4 million character samples from 120 writers. We select randomly about 92% samples from the Nakayosi database to form the training data set, 75% samples from the Kuchibue database to form the testing data set T_0, while the remaining samples from both databases are used to form a development set for tuning control parameters. By this partition, there are 704,650 samples in the training set, 229,398 in the development set, and 506,848 in the testing set, respectively. To simulate various distortions, additional 6 testing sets are generated artificially by rotating each testing sample in T_0 by $\pm10^0$, $\pm20^0$, $\pm30^0$, respectively, which are denoted as T_{10}, T_{-10}, T_{20}, T_{-20}, T_{30}, T_{-30}, respectively.

Although other options exist, for simplicity, we use an MP-based classifier to evaluate the effectiveness of the proposed affine distortion compensation approach. Again, a 512-dimensional raw feature vector [2] is extracted from each compensated sample. Then a new 192-dimensional feature vector is obtained via LDA transformation estimated from training samples of 2965 Kanji characters. A two-prototype-based classifier is trained from 192-dimensional training feature vectors by using a new minimum classification error (MCE) training approach (i.e., MCE3) we proposed recently in [4]. As for CDHMMs, the number of Gaussians per state is 4. Readers are referred to [3, 7] for other details.

3.2 Experimental Results

Let ϕ_{art} denote the artificial rotation degree, and ϕ_{est} the estimated one. Two frequency distributions of estimation errors, $\Delta\phi_{min} = \min|\phi_{art} - \phi_{est}|$, by using "Approach 2" and "Approach 3" to estimate ϕ_{est}, are tabulated in Tables 1 and 2, respectively, where the bin starts from 0^0 with a bin size of 5^0. In the last row of these Tables, the average number of orientation hypotheses is also given for each testing set. Compared to the average number of 1.89 orientations hypothesized by "Approach 1" as reported in [5], a much smaller number of 1.058 and 1.02 is achieved by "Approach 2" and "Approach 3", respectively. By comparing the results in Table 1 with that in [5], "Approach 2" achieves similar estimation errors as "Approach 1" with much smaller number of orientation hypotheses. However, an increased estimation error $\Delta\phi_{min}$ is observed for "Approach 3".

Table 3 summarizes the "Top-N" character recognition accuracies (in %) of the SP-based classifier described in section 2.1 with orientation compensation using "Approach 1", "Approach 2" and "Approach 3", respectively. It is observed that similar "Top-N" character recognition accuracies are achieved by "Approach 1" and "Approach 2", while worse performance is achieved by "Approach 3".

Table 4 compares the "Top-1" character recognition accuracies (in %) of the MP-based classifier without distortion compensation ("Baseline") with that of using the proposed affine distortion compensation based on "Approach 2" and "Approach 3" respectively. In minimax classification, a short list of top 6 candidates is used, and 3 EM iterations are performed to estimate each affine transform. It is observed that the performance of the baseline system degrades severely with an increased degree of distortions while the proposed approaches with affine distortion compensation is robust against a wide range of distortions. "Approach 3" performs slightly better than "Approach 2" for severely distorted samples (e.g. $T_{\pm20}$, $T_{\pm30}$). Unfortunately, both "Approach 2" and "Approach 3" degrade the performance for non-distorted samples

Table 3. Comparison of "Top-N" character recognition accuracies (in %) on seven testing sets achieved by an efficient SP-based pre-classifier with orientation compensation using "Approach 1", "Approach 2", and "Approach 3" respectively.

Top-N Accuracy		Testing Sets							
		T_{-30}	T_{-20}	T_{-10}	T_0	T_{10}	T_{20}	T_{30}	Ave.
Approach 1	Top-6	95.49	95.54	96.04	96.20	95.99	96.17	96.92	96.05
	Top-12	96.91	96.97	97.35	97.44	97.29	97.45	98.01	97.35
Approach 2	Top-6	95.83	95.81	95.97	96.17	95.99	96.10	96.80	96.10
	Top-12	96.91	96.93	97.25	97.31	97.10	97.25	97.83	97.23
Approach 3	Top-6	93.90	93.93	94.78	95.06	94.90	95.27	95.92	94.83
	Top-12	94.92	95.05	95.75	95.93	95.83	96.12	96.67	95.75

Table 4. Comparison of "Top-1" character recognition accuracies (in %) on 7 different testing sets by an MP-based classifier trained with MCE3 approach in [4]: without distortion compensation (Baseline) vs. with compensation using "Approach 2" and "Approach 3" respectively.

Top-1 Accuracy		Testing Sets							
		T_{-30}	T_{-20}	T_{-10}	T_0	T_{10}	T_{20}	T_{30}	Ave.
Baseline		10.36	48.02	92.66	98.36	94.39	55.99	10.97	58.68
With Compensation	Approach 2	94.67	95.77	96.39	96.47	96.24	96.06	95.79	95.91
	Approach 3	95.80	96.02	96.31	96.34	96.14	96.08	96.20	96.13

from testing set T_0.

4 Conclusion and Discussions

In this paper, we have presented an approach to compensating affine distortion of an isolated online handwritten Chinese character using combined orientation estimation and affine transform estimation with an HMM-based minimax classification approach. For handwriting samples with big rotational distortions, it is helpful to first compensate the orientation distortion, and then to use the HMM-based minimax classification approach to estimate the affine distortion against the pre-compensated sample. The effectiveness of this combined compensation approach depends highly on the accuracy of the estimated orientations. More researches are definitely needed to find an effective and efficient approach to building a handwriting recognition system which works well for both normal and distorted samples.

References

[1] Z.-L. Bai and Q. Huo, "A study on the use of 8-directional features for online handwritten Chinese character recognition," *Proc. ICDAR-2005*, pp.262-266.

[2] Z.-L. Bai and Q. Huo, "An improved approach to extracting 8-directional features for online handwritten Chinese character recognition," unpublished manuscript, 2005. See also, Z.-L. Bai, *A Study on a Goal Oriented Detection and Verification Based Approach for Image and Ink Document Analysis*, Ph.D. Thesis, The University of Hong Kong, 2006, Chapter 6.

[3] T. He, *A Study on Several Problems in Online Handwritten Chinese Character Recognition*, Ph.D. Thesis, The University of Hong Kong, 2009.

[4] T. He and Q. Huo, "A study of a new misclassification measure for minimum classification error training of prototype-based pattern classifiers," *Proc. ICPR-2008*.

[5] T. He and Q. Huo, "A character-structure-guided approach to estimating possible orientations of a rotated isolated online handwritten Chinese character," *Proc. ICDAR-2009*.

[6] Q. Huo, "A decision theoretic formulation for robust automatic speech recognition," *Pattern Recognition in Speech and Language Processing*, Wu Chou and Biing-Hwang Juang (Eds.), CRC Press, 2003, pp.81-114.

[7] Q. Huo and T. He, "A minimax classification approach to HMM-based online handwritten Chinese character recognition robust against affine distortions," *Proc. ICDAR-2007,* vol. 1, pp.43-47.

[8] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition," *IEEE Trans. on PAMI*, vol. 9, pp.149-153, 1987.

[9] C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Computer Speech and Language*, vol. 9, pp.171-185, 1995.

[10] M. Nakagawa and K. Matsumoto, "Collection of on-line handwritten Japanese character pattern databases and their analysis," *International Journal on Document Analysis and Recognition*, vol. 7, pp. 69-81, 2004.