A Statistical Matching Method in Wavelet Domain for Handwritten Character Recognition

Te-Wei Chiang
Department of Accounting
Information Systems
Chihlee Institute of Technology
Taipei County, Taiwan, R.O.C.
ctw@mail.chihlee.edu.tw

Tienwei Tsai
Department of Information Management
Chihlee Institute of Technology
Taipei County, Taiwan, R.O.C.
twt@mail.chihlee.edu.tw

Sung-Lien Kang
Library & Information Center
Chihlee Institute of Technology
Taipei County, Taiwan, R.O.C.
slkang@mail.chihlee.edu.tw

Abstract — In this paper, we present a wavelet-based classification approach to recognize the characters in the rare books transcribed by ancient calligraphers. In the training phase, the wavelet-based images of the training patterns are obtained via the wavelet transform. Then, a mask is generated for each distinct class of characters through the proposed statistical analysis method of these wavelet-based images. The mask of a class is built by finding the bits of the wavelet-based images that are reliable black. Then, we can recognize an unknown character by finding the prototype character whose mask is best fitted for the unknown character. Experimental results show the effectiveness of our approach in this application domain.

Keywords — Optical character recognition, handwriting recognition, classification, wavelet transform.

1. Introduction

In ancient Chinese a lot of significant books were transcribed by elite calligraphers. Even though the characters in these rare books were handwritten in regular forms, their integrity is greatly damaged by various imperfections. The strokes are of various thicknesses, even within the confines of a single character. Additionally, characters accompanied by spots and smears outside the theoretical limits of the characters and voids within them produce enormous variations in their shapes, even though they remain legible to human readers. Furthermore, Chinese handwriting recognition is very difficult due to three factors: the character set is very large, the structure of a Chinese character is quite complex, and many Chinese characters have similar shapes.

In order to retrieve information from these books, transforming paper documents into the contents that are accessible is inevitable. One way of digitizing the contents of documents is through human typing. However, this kind of labor-intensive approach is not only time consuming, but also expensive. For over the last decade, optical character recognition (OCR) technique [12], [13] has been introduced as a practical approach for converting paper documents into computer codes. The three best known approaches for OCR are: 1) statistical approach [2], [9], 2) structural approach [5], [7], [15] and 3) neural networks (NNs) [3], [10]. In the statistical approach the measurements that describe an object (or pattern) are treated only formally as statistical variables, neglecting their “meaning”. As for the structural approach, it attempts to recognize characters the same way we generate them – as a sequence or at least a combination of strokes, straight lines, and curves. In practice, few algorithms can consistently extract structural information from characters and most structural matching approaches, however, depend largely on the correctness of the structural information. Moreover, most of the characters in the rare books were contaminated by various noises. It makes the structural approach inapplicable. On the other hand, neural networks seem to be a popular method due to the ease of realization. Anderson et al. [1] discussed the relationship between neural networks and statistical pattern recognition. They pointed out that "neural networks are statistics for amateurs. Most NNs conceal the statistics from the user." Despite these similarities, neural networks do offer several advantages such as, unified approaches for feature extraction and classification and flexible procedures for finding good, moderately nonlinear solutions. However, its result is not as good as the basic nearest neighborhood method [16], especially in the recognition of characters with low quality. Surveys of optical recognition of handwritten characters can be found in [6], [14].

Since OCR systems are quite complex, their design has to be broken down into subproblems (such as preprocessing, feature extraction, classification and postprocessing), each of which has to be solved individually, in order to be tractable. The philosophy underlying this concept is called
"divide and conquer." For brevity, we consider the design of OCR systems in terms of two subproblems: (1) feature extraction and (2) classification. Firstly, we give a brief introduction of feature extraction. Features are functions of the measurements performed on a class of objects (or patterns) that enable that class to be distinguished from other classes in the same general category. Basically, we have to extract distinguishable and reliable features from the object being classified, and then classify the object to one of the predefined classes based on these features. Although feature design would largely affect the classification performance, it has not found a general solution in most applications. They are generally designed based on the experience, intuition, and/or cleverness of the designer. There have been many attempts at automatic feature design procedures, but few have succeeded as well as hand design of features, even in very restricted domains. Consequently, we resort to develop a feature extraction method which is general and reliable but not necessary the best. Such that the recognition system is low dependent on domain-specific knowledge and can be applied to other vision-oriented applications. On the other hand, transform type features are designed for global use. Orthogonal transform such as wavelet, Walsh, Fourier, 2-D moment, DCT, and Karhunen-Loeve are used to transform 2-D character images into new domains. Information in new domains becomes useful features. In this paper, wavelet transform is used for feature extraction.

Varieties of wavelet transformations have been applied to many research fields such as pattern analysis and recognition, signal/image denoising, data compression and computer vision due to their superiority in multiresolution analysis (MRA) and spatial-frequency localization. In this paper, wavelet transforms are elaborated to decompose an object image into four subband images with different directional details. After wavelet transformation, the object image can be transformed into wavelet-based images. Afterward, to further extract reliable features from the wavelet domain, the concept of “mask matching” [13] is employed. “Mask matching” is the most obvious technique for visual pattern recognition. Imagine the unknown image being projected through a cutout (mask) onto a photosensor array. The response will be proportional to the degree of matching. Based on this concept, we can store a mask in the computer for each distinct class of characters to be recognized, and to compare the unknown characters with the stored set to find the best matching. Suppose that the character images are two-tone – all black on white backgrounds. The matching degree can be calculated by counting their matching bits, which is known as the Hamming distance criterion. However, this kind of global Boolean mask is unreliable. It is because the bits along the edge of a character image are often subject to unpredictable variations and noises. Accordingly, we develop a mask generation method for the purpose of finding more reliable features. We superimpose a number of wavelet-based images of the “same” character and calculate the fraction of time that a given bit is black. The bits that are reliably black are served as the mask. Then, we can recognize an unknown character by finding the prototype character whose mask is best fitted for the unknown character.

This paper is organized as follows. The next section presents our classification approach. Section 3 introduces the wavelet transform and how we generate masks. Section 4 describes our statistical mask matching approach. Section 5 gives the experimental results. Finally, conclusions are drawn in Section 6.

2. The Proposed Classification Approach

In this section, we first briefly address the classification problem and introduce the framework of our approach. Then we will get into more details of our approach in the following sections. In statistical approach, a pattern is typically represented by a set of $D$ features, viewed as a $D$-dimensional feature vector. The ultimate goal of classification is to classify an unknown pattern $x$ to one of $M$ possible classes ($c_1, c_2, \ldots, c_M$).

Our experimental system is operated in two phases: training and classification. In the training phase, the feature extraction module finds the appropriate features for
representing the input patterns, and the classifier is then trained using these features. In the classification phase, the trained classifier assigns the input pattern to one of the pattern classes based on the measured results. The framework of our classification approach is illustrated in Figure 1.

Note that the preprocessing module plays the role of normalizing the samples to a certain size. Basically, mask matching method is a variant of template matching method. In the template matching method, usually, normalization of position and size is done prior to the matching process. Template matching is based on the assumption that the position of each point is known, the correspondence of the sampling point is already taken, and the similarity is measured only by the difference of the sampling “value”. Therefore, the normalization process becomes very important in order to fix the sampling position in advance. Before our classification process, we use a simple geometrical transformation to “normalize” the circumscribed rectangle about a handwritten character to make the shapes of the given characters more uniform.

After wavelet transformation, a character image is transformed into a set of wavelet coefficients. For the ease of the subsequent matching process, the wavelet coefficients are quantized to binary bits that constitute the wavelet-based image via the binarization module. The details of the wavelet transform and the mask generation method are introduced in Section 3, and our statistical mask matching approach is elaborated in Section 4.

3. Feature Extraction

Based on the requirement of reliable and general features, wavelet transform is first applied to extract statistical features.

3.1. Wavelet transform

A character image of size $N \times N$ can be decomposed into its wavelet coefficients by using Mallat’s pyramid algorithm [8]. This algorithm is focused on the computation of the following equations [11]:

$$
LL(k)(m,n) = [(LL_{row}^{(k-1)} \ast H)]_{2^k \times column} \ast H_{1 \times 2^k}
$$

(1)

$$
LH(k)(m,n) = [(LL_{row}^{(k-1)} \ast H)]_{2^k \times column} \ast G_{1 \times 2^k}
$$

(2)

$$
HL(k)(m,n) = [(LL_{row}^{(k-1)} \ast G)]_{2^k \times column} \ast H_{1 \times 2^k}
$$

(3)

$$
HH(k)(m,n) = [(LL_{row}^{(k-1)} \ast G)]_{2^k \times column} \ast G_{1 \times 2^k}
$$

(4)

Here $LL$, $LH$, $HL$, and $HH$ represent four subimages of the image being decomposed, and $LL^0$ is the original character image. $H$ and $G$ correspond to the low-pass and the high-pass filters, respectively. Expression $2 \downarrow 1(1 \downarrow 2)$ denotes sampling along column (row), and $k$ is wavelet analysis times. Figure 2 shows the decomposition (Haar wavelet) of Chinese character image “金”. Figure 3 illustrates the wavelet decomposition.

3.2. Mask generation

We know that the border bits are the most unreliable; the bits at the edge of a character image are often subject to writing and scanning noise. We can see this by superimposing a number of images of the “same” character and calculating the fraction of time that a given bit is black. What we need to find are those bits that are reliably black in certain characters. To achieve this goal, we superimpose the wavelet-based images of the training samples belonging to the same class to obtain the probability of each bit (or pixel) being black in face of that class of sample. Figure 4 shows the result of superimposing the 1st-level wavelet-based images of the training samples of character class “金”. The darkest bits in the figure represent the bits that are the most reliably black, i.e., the darker the bits are the more reliably black the bits are. We can see that the unreliable bits are along the edges. Then, the darkest bits form a mask of “金”.

![Wavelet decomposition of Chinese character image “金”. (a) The original image of size 80×80; (b) The 1st level decomposition of the image; (c) The 2nd level decomposition of the image.](image)

![Illustration of the wavelet decomposition. (a) Illustration of $LL^0$; (b) Illustration of the 1st level decomposition; (c) Illustration of the 2nd level decomposition.](image)
In the general case, we use thresholds $\alpha$ to define the meaning of “reliably black”. For example, as $\alpha = 0.8$, a bit is regarded as reliably black when the percentage of the corresponding bits of the superimposed characters is above 80%.

### 4. Statistical Mask-Matching Approach

Most commonly used optimization methods in statistical approach are based on Bayes’ theorem. Our mask-matching approach is also derived from Bayes’ theorem.

#### 4.1. Bayes classification

In statistical pattern recognition, we recognize that features may be measured with error and that some of the features are useful for identification of the class while others are not. Our goals are then to obtain useful sets of features and to use these features such that the identification is as accurate as possible. If there is an object that is to be classified on the basis of a feature $x$, into $M$ possible classes ($c_1, c_2, \ldots, c_M$), then the probability of $x$ in class $i$ when $x$ is observed can be described by $P(c_i | x)$. From the “theorem on compound probabilities” [4], we obtain

$$P(x \& y) = P(x | y) \cdot P(y) = P(y | x) \cdot P(x) \quad (5)$$

In our situation, $x$ is the feature and $y$ represents the class variable $c_i$. Substituting for $x$ and $y$ in (5), we obtain the probability that the class is $i$ when the feature $x$ is observed.

$$P(c_i | x) = \frac{P(x | c_i)P(c_i)}{P(x)} \quad (6)$$

This is Bayes’ theorem, which gives the probability of a class $i$ being present when a feature $x$ is observed, provided we know the probability of the feature being observed when the class is present, the probability of that class being present, and the probability of that feature.

#### 4.2. Measures for mask matching

Before the mask-matching process, we have to define a measure to indicate the degree of matching between a sample character and a mask.

Suppose the wavelet-based character images and the masks are of the same size ($N \times N$ bitmap). The black bits are those bits with value 1 in the bitmaps, and white bits are those with value 0. Let $N_b(p)$ be the number of black bits in bitmap $p$, and $M_b(p, q)$ be the number of black bits with the same positions in both bitmap $p$ and bitmap $q$. Then, the degree of matching between an unknown character $x$ and the mask of class $i$, $m_i$, can be defined by:

$$d(x, m^i) = \frac{M_b(x, m^i)}{N_b(x)} \quad (7)$$

**Definition 1.** If $x$ matches to the mask of class $i$ at the degree of $\alpha$, i.e., $d(x, m^i) = \alpha$. It is called $x$ $\alpha$-match the mask of class $i$, and denoted by $x^i_{\alpha}$.

#### 4.3. Statistical mask-matching

If there is an unknown character $x$ that is to be classified on the basis of its matching result into one of $M$ possible classes ($c_1, c_2, \ldots, c_M$), then the probability of $x$ in class $i$ when $x^i_{\alpha}$ is observed can be described by $P(c_i | x^i_{\alpha})$. From Bayes’ theorem, we get

$$P(c_i | x^i_{\alpha}) = \frac{P(x^i_{\alpha} | c_i)P(c_i)}{P(x^i_{\alpha})} \quad (8)$$

This equation gives the probability of a class $i$ being present when the mask of the class is matched, provided we know the probability of the mask being matched when the class is present, the probability of that class being present (called the “prior probability”), and the probability of that mask being matched.

Finally, to decide the expected class of the input pattern $x$, the following decision rules are used:

1) Rule $MD$ (Matching Degree):

$$E(x) = \arg \max_{1 \leq i \leq M} \{ d(x, m^i) \} \quad (9)$$

2) Rule $MP$ (Matching Probability):

$$E(x) = \arg \max_{1 \leq i \leq M} \{ P(c_i | x^i_{\alpha}) \} \quad (10)$$

where $\alpha = d(x, m^i)$. 

![Figure 4. The result of superimposing the 1st-level wavelet-based images of the training samples of character class “金”.](image-url)
Table 1. Recognition Rate for Each Decision Rule

<table>
<thead>
<tr>
<th>Decision Rule</th>
<th>Number of Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Top 1</td>
</tr>
<tr>
<td><strong>MD</strong></td>
<td>72.6%</td>
</tr>
<tr>
<td><strong>MP</strong></td>
<td>80.1%</td>
</tr>
</tbody>
</table>

The concept underlying rule MD is very intuitively. The class whose mask matches the input pattern most will be regarded as the expected class of the input pattern. On the other hand, rules MP selects the most likely class as the expected class of the input pattern according to the probability of the class being present when its mask is matched. We expect rule MP will outperform rule MD.

5. Experimental Results

A preliminary experiment has been made to test our approach. There are a total number of 18600 samples (about 640 categories) extracted from one of the famous handwritten rare books, Kin-Guan Bible. Each character pattern was transformed into a $48 \times 48$ bitmap pattern by means of simple size normalization. The mask for each character category is generated by superimposing the character patterns of the same category. The threshold $\alpha$ used to generate masks is set to 0.8. 1000 of the 18600 samples are used for testing. Table 1 shows the recognition rate for each decision rule. As we expected, rule MP outperforms rule MD.

6. Conclusions

This paper presents a wavelet-based statistical mask-matching approach for recognizing handwritten characters in Chinese paleography. After generating the mask for each prototype character and calculating some prior probabilities in advance, we can obtain the probability of a class being present when the mask of the class is matched in a certain degree. In our preliminary experimental results, the recognition rate is about 80 percent for a unique candidate, and 89 percent for multichoice with 10 candidates.

Since only preliminary experiment has been made to test our approach, a lot of works should be done to improve this system:

- Since features of different types complement one another in classification performance, by using features of different types simultaneously, classification accuracy could be improved.
- In order to alleviate the load of the character recognition, a coarse classification scheme needs to be involved in our system.

References


