Gray-Scale Character Image Recognition Based on Fuzzy DCT Transform Features

Xuewen Wang, Xiaoping Ding, Changsong Liu
OCR Lab, Image Processing Division, Dept. of E.E., Tsinghua Univ., P.R.China
E-mail: wxw@ocserv.ee.tsinghua.edu.cn

Abstract

We propose a new method for recognizing gray-scale text images with noise and of extraordinary low resolution. First, a fuzzy classification of pixels in the gray-scale character images is applied to form the fuzzy attributed pixel graph using integral ratio techniques. Second, DCT transform was applied on this fuzzy graph and parts of the frequency domain components are selected and compressed to produce the features for recognition. By using the modified quadratic discriminant function in the classifier, we achieve satisfactory recognition results with the average recognition rate of 96.82% on 15pixel×15pixel character images and at the speed of over 30 characters/s.

1. Introduction

This paper describes a new method for character recognition of gray-scale images. It differs from earlier works [2][3] advocating gray-scale feature selection in that it makes use of the fuzzy property of the pixels. In traditional OCR research, character images are supposed as bi-level images and the gray-scale image input from the scanner must be binarized before feature extraction. For documents in good quality and scanned with adequate resolution, the binarization will not be a serious problem. However, when the gray-scale image with noise and of low resolution is binarized, it will result in a lot of information loss and the recognition rates will decrease. Figure 1 shows an example of gray-scale character images in low resolution and their corresponding binary images.

Previous work on gray-scale image processing designed for character image recognition can be divided into two categories: (a) Use optimal thresholding methods to carefully binarize the image and extract traditional recognition features from it; (b) Extract recognition features directly from gray-scale character images without the binarization process.

Adaptive thresholding methods in the first category have been studied for a long time, but the problem remains unsolved. In many cases, the thresholded image often presents attached or damaged characters.

There have been several methods proposed in the second category. These methods extract features such as topographic features [2] or boundary features [3] directly from the gray-scale character images. But all these methods are time-consuming and do not perform well in low resolution when recognizing Chinese characters.

In the proposed method, we extract the fuzzy DCT transform features from the gray-scale images using the fuzzy properties of pixels. In Part 2 and 3 of this paper we will give detailed descriptions of this method. In Part 4, some experiments will be carried out to verify the proposed method.

Figure 1. (a) is a gray-scale image in 75dpi, and (b) is the binary image derived from (a) using morphological filters [1].

2. Concept of the fuzzy attributed pixel graph of the gray-scale character image

When we scan the document and digitize it, the obtained image will be a degraded form for the reasons described in [2]. To simplify the analysis below, we will consider only the degradation coming from the PSF of the scanner.

The PSF of a scanner is a bell shape function and can be well approximated by the function [2]:

\[ h(x,y) = c(x)c(y) , \]

where

\[ c(x) = \begin{cases} \frac{1}{\sigma} \cos^2 \left( \frac{\pi x}{2\sigma} \right) , & |x| \leq \sigma \\ 0 , & otherwise \end{cases} \]

This function has a finite support and has been widely used in optics.

The gray-scale image obtained after scanning is given by the convolution of the ideal bi-level character image with h(x,y). Here we take the vertical bar symbol as our prototype model (Fig. 2(a)):

\[ f(x,y) = \begin{cases} 1 , & |y| \leq a, |x| \leq b \\ 0 , & otherwise \end{cases} \]

where \(a \neq b\).
We can see that after the convolution, slopes are formed in the boundary areas, and there is no clearly distinction between the background and the character stroke in the gray-scale bar image (Fig. 2(b)). Moreover, the slopes will expand while $\sigma$ increase. If we binarize it as the precise image, a lot of pixels will be classified incorrectly, especially in the slope area. In low-resolution character images, $\sigma$ is relatively large contrasting with the stroke width $2a$. This kind of misclassification will degrade the subsequent feature extraction process greatly and decrease the recognition rate of the whole system badly.

![Figure 2. (a)Vertical bar symbol; (b)gray-scale image of the bar; (c) side view of gray-scale bar along x-axis.](image)

The main ideal of the proposed method is that we classify the pixels of gray-scale character images into three classes: background pixel class, character stroke pixel class and fuzzy pixel class, and extract features directly on all these pixels. As shown in Fig.2(c), the background pixel class contains the pixels belonging to the background, whose intensity is smaller than $T_0$; the character stroke pixel class contains the pixels in the strokes of the character, whose intensity is greater than $T_1(T_1>T_0)$; the pixels left are hard to determine whether they are background pixels or stroke pixels, so they are classified into fuzzy pixel class. Fig.3 shows the typical histogram of gray-scale character images in real documents and the classification of three pixel classes. In Part 3.1 we will describe the algorithms to select the $T_0$ and $T_1$ thresholds.

After classifying the pixels into three classes, the gray-scale character image $f(x,y)$ can be transformed to the fuzzy attributed pixel graph $g(x,y)$ using the function:

$$g(x,y)=\begin{cases} 
1, & f(x,y) \geq T_1 \\
\frac{f(x,y)-T_0}{T_1-T_0}, & T_0 < f(x,y) < T_1 \\
0, & f(x,y) \leq T_0 
\end{cases}$$

(1)

$g(x,y)$ indicates the fuzzy degree that the pixels belong to character stroke class. As described in Part 3.3, we will extract fuzzy DCT transform features directly from the fuzzy graph.

![Figure 3. Three pixel classes in character images](image)

3. Method for recognition using fuzzy DCT transform features

In this section, we present a method for extracting fuzzy DCT transform features and design a classifier based on these features. The method consists of four following steps.

3.1. Generation of fuzzy attributed pixel graph

First we smooth the gray-scale intensity histogram of character images, and find the background noise peak $p_b$ and the character stroke peak $p_c$ in this histogram [8]. Then we make use of Native Integral Ratio (NIR) technique [5] to select the thresholds of $T_0$ and $T_1$:

$$T_0 = p_b + \arg \max_{u=1,\ldots,n} f_1(u)$$

$$T_1 = p_c - \arg \max_{u=1,\ldots,n} f_2(u)$$

(2)

(3)

where $f_1(u), f_2(u)$ are NIR estimators of the form:

$$f_1(u) = \frac{\sum_{x_i=p_x^u}^{x_i+p_x^{u+1}} h(x_i)}{x_i=p_x^u}, f_2(u) = \frac{\sum_{x_i=p_x^u}^{x_i+p_x^{u+1}} h(x_i)}{x_i=p_x^u}$$

(4)

Note that $x_i$ denotes the gray scale intensity value and $h(x_i)$ denotes the number of pixels which have intensity $x_i$.

After that, we can generate the fuzzy attributed pixel graph $g(x,y)$ by applying Function (1).

3.2. Normalization of the fuzzy attributed pixel graph
For the convenience of DCT transform and feature extraction, we apply the linear interpolation method [6] to resize the fuzzy attributed pixel graph \( g(x,y) \) to a normalized 32\text{x}\times32\text{x} pixel graph \( g'(x,y) \). Also we can apply low-pass filter here to decrease the deformations introduced by the normalization.

### 3.3. Feature extraction and compression

We make the Discrete Cosine Transform (DCT) (5) on the normalized 32\text{x}\times32\text{x} pixel graph \( g'(x,y) \) to get the initial feature vector:

\[
F(x,y) = \alpha(x)\alpha(y) \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} g(m,n) \cos \left( \frac{(2m+1)\pi x}{2N} \right) \cos \left( \frac{(2n+1)\pi y}{2N} \right),
\]

\[
\alpha(x) = \begin{cases} \frac{1}{\sqrt{N}}, & x = 0 \\ \frac{1}{\sqrt{2N}}, & x \neq 0 \end{cases}, \quad \alpha(y) = \begin{cases} \frac{1}{\sqrt{N}}, & y = 0 \\ \frac{1}{\sqrt{2N}}, & y \neq 0 \end{cases},
\]

where \( N=32 \).

We select the low frequency domain components of \( F(x,y) \) to form the initial feature vector \( V_0 \) and normalize it to a uniform-length vector \( V_i \) using the following Equations:

\[
V_0 = [F(x,y)]_{0 \leq x,y < N/2-1}
\]

\[
V_i = V_0(i) \sqrt{\frac{\sum_{j=0}^{d-1} V_0(j)^2 - \delta}{d}}, \quad 0 \leq i < d
\]

The dimension of \( V_i \) is \( d=(N/2)\times(N/2)=256 \), it can be reduced by adopting appropriate feature compression algorithms. Moreover, feature compression can improve the generalization ability of classifiers and reduce the computational requirements of recognition. Here, the feature compression is defined as:

\[
V_c = \Phi_i V_i
\]

where \( V_c \) is a \( n \)-dimension feature vector and \( \Phi_i \) is a \( d \times n \) transform matrix \( (n \leq d) \). In Linear Discriminant Analysis (LDA) [7], which we use here, the goal is to find the transform \( \Phi \) such that \( |\Phi^T \Sigma_b \Phi|/|\Phi^T \Sigma_w \Phi| \) is maximized. Here \( \Sigma_b \) is the between-class scatter matrix and \( \Sigma_w \) is the within-class scatter matrix of character classes. It can be proved that such a transform \( \Phi \) is composed of \( n \) eigenvectors corresponding to the \( n \) largest nonzero eigenvalues of \( \Sigma_w^{-1} \Sigma_b \).

The selection of \( n \), the dimension of output feature vector, should maximize the recognition rate of the method. From our experience, the value of \( n \) is among 80–120 when applying the MQDF classifier [see Part 3.4] to recognize the printed Chinese characters.

### 3.4. Classifier design

The compressed feature vector \( V_c \) is input into a minimal-distance classifier. In general, the Quadratic Discriminant Function (QDF) which is optimal in the Bayesian sense for classification should be used. But the QDF suffers from the performance degradation for estimation errors of the covariance matrix and the requirement of much computation time and storage. Here we apply the Modified Quadratic Discriminant Function (MQDF) proposed by F. Kimura [4]:

\[
d(V_c) = \frac{1}{h^2} \left[ \frac{\|V_c - \mu_M\|}{\sum_{i=1}^{k} \frac{1}{2} \left( \frac{h^2 - \lambda_i}{\lambda_i} \right) \varphi_i^T (x - \mu_M)^T \right] + \log \left( h^{2(n-k)} \prod_{i=1}^{k} \lambda_i \right),
\]

where \( \mu_M \) is the mean vector of each class, \( \lambda_i \) and \( \varphi_i \) are the \( i \)th eigenvalue and eigenvector of \( \Sigma_M \), which is the maximum likelihood estimate of the covariance of \( V_i \); \( h^2 \) should be selected as a small constant and the value of \( k \) should be selected so that the estimation errors of \( \lambda_i \) and \( \varphi_i \) do not become too large. In our system we select \( k \) as a constant with the value of 32. Detailed description of this algorithm can be found in F. Kimura’s paper [4].

### 4. Experiments

Experiments are carried out with gray-scale character images of printed documents. There are totally 160 sets of samples, each sample contains 3755 Chinese characters. All the characters are in the same font called as “song ti”. Most of the images are 200–250 dpi, some are 75–150 dpi. For images scanned with the resolution of 75 dpi the typical size of a character is 15\text{x}\times15\text{x} pixel. We randomly select 130 sets from the above samples for training and the others for testing.

In order to verify the proposed method, another two methods have been applied in this experiment for the purpose of comparison.

**Method 1:** The author’s method.

**Method 2:** In this method, we binarize the gray-scale character images using the morphological filters [1] instead of generating fuzzy attributed pixel graphs, the steps followed are the same as in Method 1.

**Method 3:** This method performs binarization step and extracts directional-histogram features [4]. The classifier is a multi-template one using Euclidean distance. This method has been proved to be the best one in recognizing the printed Chinese character images in good quality and has been used widely in commercial products.

Results of error rates of the testing sets for these three methods are shown in Table 1. Here we calculate them in the group of resolutions. The proposed method achieves the lowest error rate among the three methods, especially when character images are in 75 dpi.

| Table 1. Character recognition error rates (%) |
|-----------------|-----------------|-----------------|
| Method 1 | Method 2 | Method 3 |
| 0.00 | 0.00 | 0.00 |

Proceedings of the International Conference on Pattern Recognition (ICPR'00)
<table>
<thead>
<tr>
<th>dpi</th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>0.02</td>
<td>0.02</td>
<td>0.26</td>
</tr>
<tr>
<td>200</td>
<td>0.03</td>
<td>0.03</td>
<td>0.51</td>
</tr>
<tr>
<td>150</td>
<td>0.10</td>
<td>1.95</td>
<td>6.23</td>
</tr>
<tr>
<td>100</td>
<td>0.36</td>
<td>0.82</td>
<td>7.06</td>
</tr>
<tr>
<td>75</td>
<td>3.18</td>
<td>6.06</td>
<td>27.99</td>
</tr>
</tbody>
</table>

Table 2. Character recognition speed (characters/s) in PC with a 350Mhz PII processor

<table>
<thead>
<tr>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>31</td>
<td>78</td>
</tr>
</tbody>
</table>

In order to test the effects of noise on the proposed method, we carry out another experiment in which a lot of real documents with artificial noises are recognized. The total character number in the documents is 11,688 and the font is the same as above. All the character images are scanned with the resolution of 300 dpi. In this experiment we generate three types of noises artificially: (1) Additive Gaussian noise with the form $N(0, \sigma^2)$; (2) Additive salt and pepper noise; (3) Multiplicative speckle noise with the mean being zero. In Method 3, the binary images are obtained from the gray-scale images by applying median filters and selecting binarizing thresholds manually (Table 3.). The recognition results are shown in Table 4.

Table 3. Examples of corrupted character images

<table>
<thead>
<tr>
<th>Noise type</th>
<th>Parameters of artificial noises</th>
<th>Gray-scale</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>No noise</td>
<td></td>
<td>功课</td>
<td>功课</td>
</tr>
<tr>
<td>Gaussian noise</td>
<td>$N(0,3)$</td>
<td>功课</td>
<td>功课</td>
</tr>
<tr>
<td>Salt &amp; pepper noise</td>
<td>5%</td>
<td>功课</td>
<td>功课</td>
</tr>
<tr>
<td>Speckle noise (mean = 0)</td>
<td>Variance is 3</td>
<td>功课</td>
<td>功课</td>
</tr>
<tr>
<td></td>
<td>Variance is 13</td>
<td>功课</td>
<td>功课</td>
</tr>
</tbody>
</table>

Table 4. Recognition rates of corrupted images (%)

<table>
<thead>
<tr>
<th>Noise type</th>
<th>Parameters of artificial noises</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>No noise</td>
<td></td>
<td>98.88</td>
<td>99.03</td>
<td>97.08</td>
</tr>
<tr>
<td>Gaussian noise</td>
<td>$N(0,3)$</td>
<td>98.35</td>
<td>97.23</td>
<td>91.34</td>
</tr>
<tr>
<td>Salt &amp; pepper noise</td>
<td>5%</td>
<td>97.00</td>
<td>91.31</td>
<td>88.57</td>
</tr>
<tr>
<td>Speckle noise</td>
<td>Variance is 3</td>
<td>95.87</td>
<td>98.49</td>
<td>95.58</td>
</tr>
<tr>
<td></td>
<td>Variance is 13</td>
<td>92.31</td>
<td>89.72</td>
<td>91.18</td>
</tr>
</tbody>
</table>

The experiment shows that the proposed method is much more robust than the other two methods when recognizing character images with noise.

5. Conclusion

In this paper we propose a new method for recognizing gray-scale text images. First, a fuzzy classification of pixels in the gray-scale character images is applied to form the fuzzy attributed pixel graph, which makes use of integral ratio techniques based on gray-scale histograms. Second, DCT transform is applied on this fuzzy graph and parts of the frequency domain components are selected and compressed to produce the features for recognition. By using the Modified Quadratic Discriminant Function in classification, this method is especially effective for recognition of images with noise and of low resolution.

Contrasting with the traditional bi-level thresholding method, the fuzzy attributed pixel graphs generated by the fuzzy classification of the pixels in gray-scale images avoid a lot of information loss. For character images of low resolution, which have relatively “spread” boundaries along the character strokes, this method will significantly improve the performance of character recognition.

Acknowledgments

This work was supported by 863 Hi-tech Plan (project 863-306-ZT03-03-1) & National Natural Science Foundation of China (project 69972024).

References