Progressive Pattern Matching Approach Using Discrete Cosine Transform

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

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

Yi-Chen Lin
Department of Business Administration
Chihlee Institute of Technology
Taipei, Taiwan, R.O.C.
ychlin@mail.chihlee.edu.tw

***Abstract**-The discrete cosine transform (DCT) is a well-known technology used in image compression. It can decompose a signal into N orthogonal real base vectors. Due to its energy compacting property, the features of a test character can be extracted progressively according to their importance such that the potential candidates which are not well-fitted for the test character can be pruned at earlier steps. Therefore, we are motivated to develop a progressive matching approach based on DCT for recognizing of handwritten characters in Chinese paleography. Experimental results show that our approach performs well in this application domain.*

Keywords: Optical character recognition, handwriting recognition, classification, discrete cosine transform.

1. Introduction

In ancient Chinese a lot of significant books were transcribed by elite calligraphers. 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 [7] has been introduced as a practical approach for converting paper documents into computer codes.

In the literature, some methods had proposed for OCR, such as methods based on neural networks (NNs) [2], [4], graph-matching methods [3], and methods based on hidden Markov model [6], etc. Due to the ease of realization, neural networks seem to be the most popular method; however, its result is not as good as the basic nearest neighborhood method [12], especially in the recognition of characters with low quality. Even though graph-matching method has strong theoretically background, its effect is not satisfactory because of the difficulty in combination with feature extraction methods. Methods based on hidden Markov model have been employed to cope with the recognition of connected characters (such as English characters). However, Chinese handwriting recognition is much more difficult than alphabetic characters. The difficulty is due to three factors: the character set is very large, the structure of a Chinese character is much more complex than that of an alphabetic character, and many Chinese characters have similar shapes.

To further improve the efficiency of the aforementioned technologies, we can introduce the algorithms used in the compression technologies. The rapid growth of digital imaging applications, including desktop publishing, multimedia, teleconferencing, and high-definition television (HDTV) has increased the need for effective and standardized image compression techniques. There are some emerging standards, such as JPEG, for compression of still images [10]; MPEG, for compression of motion video [8]; and CCITT H.26x, for compression of video telephony and teleconferencing. All three of these standards employ a basic technique known as the discrete cosine transform (DCT). Developed by Ahmed et al. [1], the DCT is a close relative of the discrete Fourier transform (DFT) which transforms a signal or image from the spatial domain to the frequency domain. The DCT is a technique for converting a signal into elementary frequency components. Each DCT uses N orthogonal real basis vectors whose components are cosines. The DCT approach has an excellent energy compaction property and requires only real operations in transformation process. The DCT helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality).

The DCT is less sensitive to shift, rotate and scale, which would largely affect the performance of most template matching methods. 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". In most of the template matching methods, normalization of position and size is done prior to the matching process [5], [9]. Therefore, the performance of normalization is an important factor to the fixing the sampling position, which is also an important issue in the matching process. In contrast, the perform-

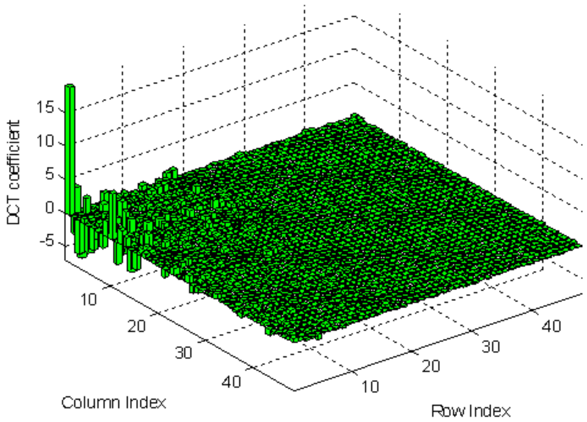


Figure 1. The DCT coefficients of the character image of "佛".

ance of normalization is not an important issue for the matching method based on the DCT.

Based on above observations, we were motivated to devise a progressive matching approach based on DCT for recognizing handwritten characters in Chinese paleography. In our approach the DCT is applied to transform the character images into elementary frequency components. Due to the energy compacting property of DCT, much of the signal energy has a tendency to lie at low frequencies. Therefore, to recognize an unknown character, only the candidates whose DCT coefficients of low frequencies well-match those of the unknown character need to be further examined. Such that the matching process can be conducted progressively from low frequencies to high frequencies and terminated if the accumulated matching cost exceeds a predefined threshold. So we can recognize an unknown character by finding the prototype character whose DCT coefficients are best fitted for the unknown character. Experimental results show that our approach performs well in this application domain.

This paper is organized as follows. The next section introduces the DCT. Section 3 describes our progressive matching approach based on the DCT. Section 4 presents experimental results. Finally, conclusions are drawn in Section 5.

2. Discrete Cosine Transform

The DCT is introduced in this section. Its most important feature is its superior energy compacting property. On applying DCT, a frequency spectrum (or the DCT coefficients) $C(u, v)$ of an $N\times N$ image represented by $x(i, j)$ for i, j=0, 1, …, N-1 can be defined as

$$C(u,v) = \frac{2}{N}\alpha(u)\alpha(v)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}x(i,j) \times \cos\left(\frac{(2i+1)u\pi}{2N}\right)\cos\left(\frac{(2j+1)v\pi}{2N}\right), \quad (1)$$

where

$$\alpha(w) = \begin{cases} \sqrt{1/2} & for\ w = 0, \\ 1 & otherwise. \end{cases}$$

The corresponding inverse discrete cosine transform (IDCT) is defined as

$$x(i,j) = \frac{2}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\alpha(u)\alpha(v)C(u,v) \times \cos\left(\frac{(2i+1)u\pi}{2N}\right)\cos\left(\frac{(2j+1)v\pi}{2N}\right). \quad (2)$$

We notice that DCT is a unitary transform, which has the energy preservation property, i.e.,

$$E = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left(x(i,j)\right)^2 = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left(C(u,v)\right)^2, \quad (3)$$

where E is the signal energy.

For most images, much of the signal energy lies at low frequencies; these appear in the upper left corner of the DCT. The lower right values represent higher frequencies, and are often small - small enough to be neglected with little visible distortion. Figure 1 shows the DCT coefficients of a character image ("佛") of size 48×48. The number of coefficients is equal to the number of pixels in the character image. In our experiments, we found that about 90% signal energy appears in the upper left corner of size 10×10 for the DCT coefficients of size 48×48.

3. Progressive Matching Method

Before going into the details of our progressive matching method, we first briefly present the problem and related symbols used in this paper. Our goal is to classify an unknown character x to one of M possible classes (c_1, c_2,…, c_M). In order to apply the template matching method, the database that consists of the templates for each class must be established in advance. Therefore, in the training phase, a set of training character images of size $N\times N$ is collected and transformed into DCT coefficients. Then the average DCT coefficients of size $N\times N$ is obtained for each class, assuming there is at least one training character for each class. Such that M sets of average DCT coefficients are obtained and served as the templates for each class.

Based on the observation that the most significant DCT coefficients lie in its low frequency area, therefore, the matching process can be performed step by step by adding more and more coefficients from low frequency area to high frequency area. Thus, we can develop a matching algorithm that can progressively measure the distance (or dissimilarity) between the unknown character x and template T_i.

3.1. Matching criterion

Firstly, we have to define a measure to indicate the degree of matching between a test character and a template. Traditionally, we can apply the sum of absolute differences (SAD) as the matching criterion to avoid multiplications. To exploit the energy preservation property of DCT, we use the sum of squared differences (SSD) instead, which can be implemented efficiently using a look-up table to calculate the square. Assume that $C_x(u, v)$ and $C_{Ti}(u, v)$ represent the DCT coefficients of the test character x and template T_i, respectively. Then the SSD between x and T_i can be defined as

$$SSD(x,T_i) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left(C_x(u,v) - C_{T_i}(u,v)\right)^2. \quad (4)$$

For the ease of progressive matching, we also define the SSD between x and T_i under the upper left block of size $n \times n$ as

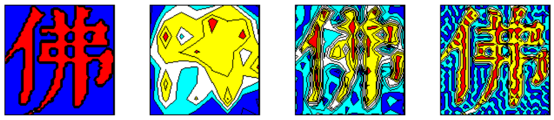
$$SSD(x,T_i,n) = \sum_{u=0}^{n-1}\sum_{v=0}^{n-1}\left(C_x(u,v) - C_{T_i}(u,v)\right)^2, \quad (5)$$

assuming the upper left $n \times n$ DCT coefficients of x and T_i are correspond to a block of size $n \times n$.

3.2. Description of the algorithm

The matching of x and T_i can be decomposed into K iterations, each of which corresponds to the matching under the block of size $n_k \times n_k$. After the kth iteration, the block size is enlarged from $n_k \times n_k$ to $n_{k+1} \times n_{k+1}$ ($n_{k+1} = n_k + \delta$). At the first iteration (k=1), the lowest frequency coefficients are measured; then the second lowest frequency coefficients are measured at the second iteration (k=2), and so on, until one of the stop criterions is satisfied. The stop criterions used in our approach are twofold: to preserve enough signal energy in the block and to reject unqualified templates at earlier steps. For example, since about 90% signal energy appears in the block of size 10×10 for an image of size 48×48, the matching process will be stopped as the block size reaches to 10 if 90% signal energy is used as the threshold for stop criterion. Therefore, the maximum number of iterations for a matching depends on the initial block size, the increment δ and the stop criterion of energy preservation. On the other hand, to reject the unqualified templates in the early iterations, statistical tools are applied to calculate the theoretical sound thresholds for the SSD under different block size. From the Chebyshev rule [11], we know that the percentage of observations contained within distances of k standard deviations around the mean must be at least $(1-1/k^2)\times 100\%$. For example, 97.2% is guaranteed within the distances of 6 standard deviations. Assume that θ_n represents the threshold for the block of size $n \times n$. To obtain the threshold θ_n, all of the SSDs between the test characters and their target templates under the block of size $n \times n$ are gathered. Then the mean μ_n and the standard deviation σ_n of SSD under the block of $n \times n$ can be calculated. Thus, θ_n can be obtained by

$$\theta_n = \mu_n + k \times \sigma_n. \quad (6)$$



(a) (b) (c) (d)

Figure 2. (a) The original image of size 48×48; (b) The reconstructed image of size 8×8; (c) The reconstructed image of size 16×16; (d) The reconstructed image of size 32×32.

The input of the progressive matching algorithm is the test character x and a template T_i. The output of the algorithm is either the distance between x and T_i or "Reject" the template. If T_i is accepted as a candidate for the test character, then output the distance between x and T_i, $SSD(x, T_i, N')$, under the block of size $N' \times N'$, assuming N' is the minimum block size that satisfies the aforementioned stop criterion. Otherwise, output "Reject the template". The algorithm is described as follows.

Step 1: Set the initial block size to n.

Step 2: Compute $SSD(x, T_i, n)$.

Step 3: If $SSD(x, T_i, n) > \theta_n$, then output "Reject" and stop.

Step 4: If $n \geq N'$, then output $SSD(x, T_i, N')$ and stop.

Step 5: Set $n = n + \delta$ and go to Step 2.

This algorithm is very efficient because the results for calculating coefficients of smaller block can be used for calculating the following coefficients of larger block. Finally, to decide the expected class

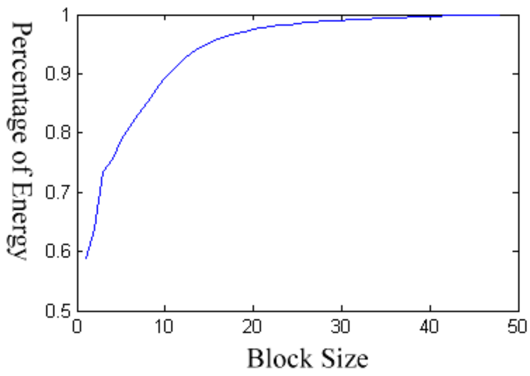


Figure 3. The percentage of energy preserved as the block size is changed.

of the test character x, the class of the template with minimum $SSD(x, T_i, N')$ is selected.

The underlying philosophy of the progressive matching algorithm is the concept of multi-resolution. Adding more DCT coefficients usually imply increasing the resolution level of an image. If current resolution is not high enough to distinguish one character from the others, we have to raise the level of resolution such that the discrimination capability can also be improved. Figure 2 shows the images reconstructed from different block size via IDCT. From this figure, we conclude that the bigger the block size, the higher the resolution of the reconstructed image.

4. Experimental results

A preliminary experiment has been made to test our approach. There are a total number of 6000 samples (about 500 categories) extracted from one of the famous handwritten rare books, Kin-Guan (金剛) bible. Each character image was transformed into a 48×48 bitmap. 5000 of the 6000 samples are used for training and the others are used for testing. Figure 3 shows the percentage of energy preserved as the block size of the upper left DCT coefficients is changed. From this figure, it can be seen that about 84% signal energy is preserved in the block of size 8×8 and 89.3% signal energy is preserved in the block of size 10×10.

Table 1 shows the recognition rate of test samples under different block size. We can find that the recognition rate cannot be improved further as the block size is larger than 8×8. This evidence supports the hypothesis that if the compacted DCT coefficients preserved enough energy, they possess enough information for classification at the same time. Through the energy compacting property of DCT, the number of features for a character can be reduced from 48×48 to 8×8 without sacrificing the recognition rate. Therefore, the parameter used for stop criterion, N', is set to 8. Besides, the initial block

Table 1. The recognition rates under different block size

Block Size	Recognition Rate
4×4	77.01%
6×6	89.42%
8×8	91.97%
10×10	89.42%
24×24	90.88%
32×32	91.24%
48×48	91.24%

Table 2. Average percentage of remained candidates after each iteration of matching

Block Size	Percentage of Remained Candidates
4×4	56.91%
6×6	39.90%
8×8	28.82%

size and the increment δ are set to 4 and 2, respectively. Thus, during the progressive matching process, the block is enlarged from 4×4 to 8×8 (i.e. 4×4, 6×6 and 8×8), and hence the maximum number of iterations is 3.

As to the threshold θ_n, 6 standard deviations were applied (see EQ. (6)) for θ_4 to guarantee that at least 97.2% target templates were remained after the first round pruning, and 5 standard deviations were applied for θ_5 to guarantee that at least 96% target templates were remained after the second round pruning and 4 standard deviations were applied for θ_4 to guarantee that at least 93.75% target templates were remained after the third round pruning. Table 2 shows the average percentage of remained candidates after each iteration of progressive matching. It can be seen that most of the candidates are pruned in the early iteration and only 28.82% candidates need to be considered in the final decision process. The recognition rate of progressive matching is the same as that of the direct matching under the block of size 8×8, i.e. 91.97%. In other words, the number of squared differences needs to be calculated is reduced to 60.24% ([(1-0.5691)×4×4 + (0.5691-0.399)×6×6 + 0.399×8×8]/8×8) without sacrificing the recognition rate.

5. Conclusions

This paper presents a progressive matching approach based on DCT for recognizing handwritten characters in Chinese paleography. Due to the energy compacting property of DCT, the features of a test character can be extracted progressively according to their importance and hence the potential candidates which are not well-fitted for the test character can be pruned at earlier steps. The advantages of our approach include:

- Through the energy compacting property of DCT, the number of features for a character is reduced from 48×48 to 8×8.

- Through the progressive matching algorithm, the number of squared differences needs to be calculated is reduced to 60.24% without sacrificing the recognition rate.

Since only preliminary experiment has been made to test our approach, a lot of works should be done to improve this system. For example, since features of different types complement one another in classification performance, by using different types of features simultaneously, classification accuracy could be improved.

References

[1] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform", *IEEE Trans. on Comput.*, vol. C-23, pp. 90-93, 1974.

[2] J. P. Drouhard, R. Sabourin and M. Godbout, "A neural network approach to off-line signature verification using directional PDF," *Pattern Recognition*, Vol. 29, No. 3, pp. 415-424, 1996.

[3] A. J. Hsieh, Kuo-Chin Fan, and Tzu-I Fan, "Bipartite Weighted Matching for On-line Handwritten Chinese Character Recognition," *Pattern Recognition*, Vol. 28, no. 2, pp. 143-151, 1995.

[4] S. W. Lee, "Off-line recognition of totally unconstrained handwritten numerals using multi-layer cluster neural network," *IEEE Trans. Patt. Anal. Machine Intell.*, Vol. PAMI-18, no.6, pp. 648-652, 1996.

[5] S. W. Lee and J. S. Park, "Nonlinear Shape Normalization Methods for the Recognition of Large-Set Handwritten Characters," *Pattern Recognition* , vol. 27, no. 7, pp. 859-902, 1994.

[6] M. Mohamed and P. Gader, "Handwritten word recognition using segmentation-free hidden Markov modeling and segmentation-based dynamic programming techniques," *IEEE Trans. Patt. Anal. Machine Intell.*, Vol PAMI-18, no.5, pp. 548-554, 1996.

[7] M. Nadler and E. P. Smith, "Pattern Recognition Engineering," Wesley Interscience, 1993.

[8] A. Puri, "Video coding using the MPEG-1 compression standard", Society for Information Display Digest of Technical Papers, vol. 23, pp. 123-126, 1992.

[9] T. Wakahara, K. Odaka, "Adaptive Normalization of Handwritten Characters Using Global/Local Affine Transformation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no.12, pp. 1332-1341, 1998.

[10] G. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp.30-44, 1991.

[11] R. E. Walpole and R. H. Myers, "Probability and statistics for engineers and scientists," Macmillan Publishing Company, 1985.

[12] Hong Yan, "Comparision of Multilayer Neural Network and Nearest Neighbor Classifiers for Handwritten Digit Recognition," *Int. J. of Neural Systems*, vol. 6, no.4, pp. 417-423, 1995.