Coarse Classification via Discrete Cosine Transform and Quantization

Te-Wei Chiang, Department of Accounting Information Systems, Chihlee Institute of Technology
Tienwei Tsai, Department of Information Management, Chihlee Institute of Technology
Sung-Lien Kang, Library & Information Center, Chihlee Institute of Technology

ABSTRACT

In this paper a novel coarse classification scheme is proposed to speed up the classification process. To develop a coarse classification scheme which is low dependent on domain-specific knowledge, 2-D discrete cosine transform (DCT) is employed as feature extraction method for vision-based applications. Then, quantization method is applied to partition the feature space into a finite number of grids, each of which corresponds to a grid code (GC). In the training phase, the information about the grids that ever contain a training sample belonging to a particular class is gathered. In the test phase, on classifying an unknown object, a reduced set of candidate classes can be retrieved from the corresponding GC. In this way, a large number of improbable candidates can be eliminated, and only few surviving candidates need to be further examined in the subsequent fine classification process. This scheme can significantly reduce the classification time while maintaining the classification accuracy. Experiments were conducted for recognizing handwritten characters in Chinese palaeography and showed that our approach performs well in this application domain.

Keyword: Classification, Optical Character Recognition, Discrete Cosine Transform

1. Introduction

Classification of objects (patterns) into a number of predefined classes has been extensively studied in wide variety of applications such as character recognition, speech recognition, and face recognition. These applications often involve hundreds or thousands of classes. To alleviate the burden of classification process, the process is usually divided into two stages: the coarse classification (also called preclassification) process and the fine classification process. To classify an unknown object, firstly, the coarse classification is employed to reduce the large set of candidate objects to a smaller one. Then, the fine classification is applied to identify the class of the object. Basically, coarse classification should meet three requirements:

- high accuracy rate,
- high reduction capacity, and
- quick classification time.

Even though the three factors are mutually conflicted, a coarse classification algorithm should compromise among the three factors. The goal of this paper is to develop a general coarse classification scheme that can be applied to most of the vision-oriented applications such as optical character recognition (OCR), face recognition and image retrieval, etc. Before going into the details of the coarse classification problem, we first give a brief survey of the classification problem.
Generally speaking, we have to extract useful features from the objects being classified before the classification process. Thus, we may consider the design of classification systems in terms of two subproblems: (1) feature extraction and (2) classification. Firstly, we briefly introduce the classification problem. There are two distinct philosophies of classification in general, known as “statistical” [1], [2] and “structural” [3]. In the statistical approach the measurements that describe an object are treated only formally as statistical variables, neglecting their “meaning”. In other words, one can regard the object as a set of statistics extracted from certain measurements made on the object. Structural-based approach regard objects as compositions of structural units, usually called primitives. It works well when the primitives can be found easily. Compared with the structural approach, the statistical approach consists of the following advantages:

- fixed-length vector,
- robustness for noisy patterns,
- less ambiguities, and
- easy to implement.

Since our main goal is to develop a general coarse classification scheme for vision-oriented applications (such as OCR, face recognition, image retrieval, etc.), the statistical approach is more suitable than the structural approach. Therefore, in the remainder of this paper we shall focus on the statistical approach.

The weaknesses of statistical approaches are twofold. First, the statistical measures smear or average the parameters they are computing in the region over which they are computed. They are therefore fairly insensitive to boundaries internal to a region being analyzed. Second, since the statistical features are usually of high dimensionality, reduction of dimensionality becomes an issue in statistical approaches. The first weakness is insignificant because the coarse classification is our main concern and the discriminating ability is not so critical. As for the second weakness, it can be overcome via the energy compacting property of DCT and the quantization technique.

We now turn to a brief introduction of feature extraction. Features are functions of the measurements performed on a class of objects that enable that class to be distinguished from other classes in the same general category. Before the classification process, useful features must be extracted from the objects. Although feature design would largely affect the classification performance, it has not found a general solution in most applications. They are generally designed by hand, using 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. However, the discrimination ability of the features for coarse classification, which is our main concern, is not as critical as that of the features for fine classification. Therefore, instead of designing the best features for a certain application area, we attempt to find general features that can be applied properly to most application areas.

The purpose of this paper is to design a general coarse classification scheme, which is low dependent on domain-specific knowledge. To achieve this goal, we need not only reliable and general features in the feature extraction stage, but also general classification method in the coarse classification stage. Based on previous observations, we are motivated to apply the statistical approach and a technique which is widely used in image compression, known as discrete cosine transform (DCT) [4], for coarse classification. The DCT helps separate an image into parts of differing importance with respect to the image's visual quality. Due to the energy compacting property of DCT, much of the signal energy has a tendency to lie at low frequencies. Therefore, in our approach only the candidates whose low frequency DCT
coefficients close to those of the object being classified will be accepted in the coarse classification process, such that the dimension of features can be largely reduced.

In this paper, we will focus on the topic of improving the efficiency of the classification process. Reliable features of low dimensionality will be extracted for coarse classification. Using these features to perform coarse classification can eliminate a large number of improbable candidates in the early stage and therefore reduce the burden of the subsequent fine classification process. The remainder of this paper is organized as follows. Section 2 introduces the feature extraction subproblem and how to extract and quantize DCT features. Section 3 explains our coarse classification scheme. Section 4 gives experimental results. Finally, conclusions are drawn in Section 5.

2. Feature Extraction and Quantization

Based on the requirement of reliable and general features in the coarse classification process, we are motivated to apply DCT to extract statistical features.

2.1. Discrete Cosine Transform

Developed by Ahmed et al. [4], the DCT is a technique for separating the image into parts (or spectral sub-bands) of differing importance (with respect to the image’s visual quality). It 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.

On applying 2-D DCT, a frequency spectrum (or the 2-D DCT coefficients) $F(u, v)$ of an $N \times N$ image represented by $x(i, j)$ for $i, j=0, 1, \ldots, N-1$ can be defined as

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

(1)

where

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

In Eq. (1), the coefficients with small $u$ and $v$ correspond to low frequency components; on the other hand, the ones with large $u$ or $v$ correspond to high frequency components. For most images, much of the signal energy lies at low frequencies; the high frequency coefficients are often small - small enough to be neglected with little visible distortion. Therefore, DCT has superior energy compacting property. Figure 1 shows the 2-D 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. From Figure 1 it can be seen that much of the signal energy appears in the upper left corner of the 2-D DCT, which corresponds to the low frequency area of the 2-D DCT. Based on above observations, we were motivated to devise a cluster-based coarse classification scheme using low frequency 2-D DCT coefficients as discriminating features. In the next section we shall introduce our coarse classification scheme.
2.2. Quantization

Since only abstract features are needed for coarse classification, the quantization technique is applied to obtain a reduced representation of the feature set and render the features more suitable for the coarse classification.

In our approach the 2-D DCT coefficient $F(u, v)$ is quantized to $F'(u, v)$ according to the following equation:

$$F'(u, v) = \left\lfloor \frac{F(u, v)}{Q} \right\rfloor.$$

Assume that the quantized DCT coefficient $F'(u, v)$ belongs to one of the following quantized values: 0, 1, .., $q$. In other words, the span of $F(u, v)$ is partitioned into $(q+1)$ intervals, each of which corresponds to a quantized value. Figure 2 illustrates the quantization process. Figure 2(a) is a part of 2-D DCT coefficients extracted from a character image. Figure 2(b) is the coefficients after quantization ($Q=4$). It can be seen that most of the high frequency
coefficients are quantized to zero and only the most significant coefficients will be retained. Thus, dimension of the feature vector can be reduced after quantization.

3. Statistical Coarse Classification

The ultimate goal of classification is to classify an unknown pattern \( x \) to one of \( M \) possible classes \((c_1, c_2, \ldots, c_M)\). Basically, the statistical classification system is operated in two modes: training (learning) and classification (testing). In the training mode, the feature extraction module finds the appropriate features for representing the input patterns and the classifier is trained to partition the feature space. In the classification mode, the trained classifier assigns the input pattern to one of the pattern classes under consideration based on the measured features. To evaluate the performance of a classifier, the holdout method [5] can be applied. In the holdout method, the given data are randomly partitioned into two independent sets, a training set and a test set. The training set is used to derive the classifier, whose accuracy is estimated with the test set.

3.1. Proposed coarse classification scheme

Our coarse classification scheme is developed as follows. In the training mode, the features of each training sample are first extracted by DCT and quantized. Then the most significant \( D \) quantized DCT features of each training sample are transformed to a code, called grid code (GC), which corresponds to a grid of feature space partitioned by the quantization method. Obviously, the training samples with the same GC are similar and can be classified into a coarse class. This characteristic implies that: if an unknown sample has the same GC of a training sample, it may belong to the same class the training sample belongs to. Therefore, the information about all possible GCs is gathered in the training mode. Each code may correspond to many or no classes. In the classification mode, the classes corresponding to the same (or almost the same) GC of the test sample are chosen as the candidates of the test sample. Our model for coarse classification is shown in Figure 3. The preprocessing module plays the role of normalizing the samples to a certain size. The feature extraction and quantization modules have already described in Section 2. The other modules will be introduced in the following subsections.

![Figure 3. Model for coarse classification.](image-url)
3.2. Grid code transformation

After the quantization process, the most significant $D$ quantized DCT features of sample $O_i$ are obtained, say $[q_{i1}, q_{i2}, \ldots, q_{iD}]$. In this quantized feature vector, $q_{i1}$ is the most important feature and $q_{i2}$ is the second most important one, and so on. In our approach the significance of each DCT coefficient is decided according to the zigzag order as illustrated in Figure 4. It is based on the energy compacting property that low-frequency DCT coefficients usually are more important than high-frequency DCT coefficients.

Because the value of $q_{ij}$ may be negative, for the ease of operation, we transform $q_{ij}$ to positive integer $d_{ij}$ by adding a number, say $k_j$, to $q_{ij}$. $k_j$ is obtained from the statistics of the $j$-th component of the quantized feature vector of all training samples. The value of $k_j$ is set to transform the smallest value of the $j$-th feature to zero. In this way, object $O_i$ can be transformed to a $D$-digit GC, $d_1d_2\ldots d_D$. This process is what we call the grid code transformation (GCT).

From another point of view, the GCT is applied for the purpose of grouping the patterns with high similarity into the same code. This process, therefore, can be regarded as a coarse classification process.

![Figure 4. Illustration of extracting the 2-D DCT coefficients in zigzag order.](image)

3.3. Grid code sorting and elimination

After the GCT, we obtain a list of triplets $(T_i, C_i, GC_i)$, where $T_i$ is the ID of a training sample, $C_i$ is the Class ID the training sample belongs to, and $GC_i$ is the grid code of the training sample. Each training sample is associated with a triplet. Then the list is sorted according to the GC. Such that we can efficiently find the classes whose training samples belong to the GC being searched. Consequently, given the GC of a test sample, we can get a list of candidate classes of the same (or almost the same) GC for the test sample.

To further improve the efficiency of the searching process, redundant information must be removed. Redundancy occurs as the training samples belonging to the same class have the same GC. This redundancy can be eliminated by establishing an abstract lookup table that only contains the information about the GCs and their corresponding classes. Then, given a GC, this table can tell the relevant classes very quickly by binary search.
To sum up, the information about the classes within each GC is gathered in the training phase. In the test phase, on classifying a test sample, a reduced set of candidate classes can be retrieved from the lookup table according to the GC of the test sample.

4. Experimental Results: A Case Study of OCR

In our application, the objects to be classified are handwritten characters in Chinese paleography. In order to retrieve information from these books, transforming paper documents into the contents that are accessible is inevitable. Optical character recognition (OCR) is an essential process in converting paper documents into computer codes for digital libraries. Since most of the characters in these rare books were contaminated by various noises, it is a challenge to achieve a high recognition rate.

A preliminary experiment had been made to test our approach. There are a total number of 18600 samples (640 classes) extracted from one of the famous handwritten rare books, Kin-Guan (金剛) bible. Each character image was transformed into a 48×48 bitmap. 1000 of the 18600 samples were used for testing and the others were used for training. The most significant D DCT coefficients were quantized and transformed to a GC for each sample. Table 1 shows the reduction rate and accuracy rate of the test samples under different value of \( D \). It can be seen that the reduction rate increases as the size of GC, \( D \), increases. However, it appears to be a tradeoff that good reduction rate usually sacrifices its accuracy rate at the same time.

<table>
<thead>
<tr>
<th>( D )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>R.R.</td>
<td>0.402</td>
<td>0.500</td>
<td>0.512</td>
<td>0.592</td>
<td>0.593</td>
<td>0.732</td>
</tr>
<tr>
<td>A.R.</td>
<td>0.993</td>
<td>0.984</td>
<td>0.983</td>
<td>0.977</td>
<td>0.977</td>
<td>0.960</td>
</tr>
</tbody>
</table>

Table 1. Reduction and accuracy rate using our coarse classification scheme.


5. Conclusion

This paper presents a coarse classification scheme based on DCT and quantization. Due to the energy compacting property of DCT, the most significant features of a pattern can be extracted and quantized for the generation of the grid codes. Hence the potential candidates which are similar to the test pattern can be efficiently found by searching the training patterns whose grid codes are similar to that of the test pattern.

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 features of different types simultaneously, classification accuracy could be improved.
References


