COMPRESSED DOMAIN FEATURE TRANSFORMATION USING EVOLUTIONARY STRATEGIES FOR IMAGE CLASSIFICATION

Chun-Ip Chiu, Hau-San Wong and Horace H-S Ip

Department of Computer Science
City University of Hong Kong
83 Tat Chee Av., Kowloon Tong, Hong Kong.
chunip@cs.cityu.edu.hk,{cshswong,cship}@cityu.edu.hk

ABSTRACT

Recently, a number of approaches have been proposed which use compressed domain features for image retrieval and classification. While the main motivation of these approaches are to improve processing efficiency and reduce computational requirement, we propose a method which also aims at enhancing the content characterization capabilities of the compressed domain features in addition to efficiency improvement. In this work, we model the compressed domain features values as random variables and approximate their associated probability mass functions as histograms. We then transform these histograms in such a way that the resulting classification rate based on these transformed histograms would be improved. With a large number of possible transformations, we adopt Evolutionary Strategy (ES) to search for the optimal one. Experiments show that our proposed approach is able to obtain better classification rate while the efficiency advantage of using compressed domain features is retained.

1. INTRODUCTION

Nowadays, with the universal adoption of international compression standards like JPEG and MPEG, etc., multimedia information is usually stored in compressed format. The main benefit from the adoption of such standards is that the resulting storage size can be greatly reduced. However, with the growth of digital photography, multimedia system and the Internet, sizes of multimedia collections have been increased dramatically. As a result, compressed domain information processing has become one of the most intensive research areas. One of the most important issues is how to effectively index and retrieve compressed multimedia information as the problem has not been adequately addressed when these standards are developed.

Over the years, for content-based image retrieval and classification, feature extraction techniques are mostly developed in the spatial domain where images are processed in a pixel-by-pixel format. Therefore, to apply those techniques for compressed images, full decompression is necessary which leads to expensive computational requirements. To reduce the amount of computation in this process, as typified by inverse DCT stage of the popular JPEG DCT-based compression standards [1], there are recent attempts to perform indexing by extracting features directly in the compressed domain. In [2], the authors proposed extracting statistical parameters directly in the DCT domain to characterize the texture and shape features. They selected the DC coefficient and a subset of AC coefficients to form a feature histogram for image retrieval. In [3], DCT coefficients are reordered into different sub-bands in a way similar to wavelet multi-resolution analysis to capture the texture features. Both approaches are able to demonstrate their efficiency improvements by avoiding full decompression. While the main objective of most of these approaches is to identify and exploit features directly in the compressed domain for efficiency improvement rather than any anticipated enhancement of content characterization capability, our approach, in contrast, directly addresses the content characterization issue by explicitly searching for the optimal transformation for the compressed domain features for classification improvement based on evolutionary computation techniques

2. COMPRESSED DOMAIN FEATURES

From recent researches, it is demonstrated that a selected set of DCT coefficients is indicative of the underlying content of the original image to a certain extent and thus can be used for image classification and retrieval. In

0-7803-8554-3/04/$20.00 ©2004 IEEE.

content-based image classification and retrieval, one of the most critical issues is how to characterize the underlying content of an image. Different types of histograms have been used to summarize the statistics of extracted features like color and texture. In this section, we will describe how these features are represented.

2.1. Color Histogram

Typically, in spatial domain, a color histogram is constructed based on the occurrence frequencies of particular color intensities by scanning the image pixel-by-pixel. Similarly, in the compressed domain, we can construct a DCT coefficient histogram by directly accessing the compressed domain coefficients. For a typical DCT-based JPEG compressed image, the upper left coefficient of each block -- DC coefficient actually represents the average intensity of the block which can characterize the brightness and color contents. In this work, we have adopted the luminance and chrominance features which are characterized by the DC coefficients in the Y luminance block and the Cb and Cr chrominance blocks. To construct a compressed domain feature histogram to characterize the color contents, we can simply count the occurrence frequency of a particular value of the DC coefficients in all the blocks of an image. In other words, the chosen DCT coefficient is modeled as a random variable x which takes values in the set $X = \{x_1, \ldots, x_u, \ldots, x_U\}$ and its associated histogram is an approximation of the variable's associated probability mass function $P(x_u) = \Pr\text{ob}(x = x_u)$.

2.2. Edge Orientation Histogram

In [4], the authors proposed a fast algorithm for extracting edge information directly in the DCT domain. Essentially, besides the DC coefficient, different AC coefficients in a DCT block represent the energy in certain frequency sub-bands. This information can be used to detect and extract edge information within a block. Specifically, the authors used an ideal edge model to facilitate edge parameter estimation such as its orientation, its offset and strength. Although the extracted edge is coarse, it is adequate for classification purpose since we only need to characterize the gross visual features of each image class. In this work, we have mainly adopted the edge orientation parameter for image classification purpose. We used the following metric proposed in [4] to acquire the edge orientation information $\tan\theta = \left(\sum_{p=1}^{7} F_{0p}\right) / \left(\sum_{q=1}^{7} F_{q0}\right)$ where F_{pq} stands for the pq-th coefficient in the 8×8 DCT block. Similar to color feature processing, we model the edge orientation value $\tan\theta$ as another random variable y which takes values in the set $Y = \{y_1, \ldots, y_v, \ldots, y_V\}$ and construct its histogram as an approximation of its probability mass function $P(y_v) = \Pr\text{ob}(y = y_v)$.

3. FEATURE TRANSFORMATION

If the histogram is used for classification, then it is apparent that different transformations on the random variable will reshape the histogram in distinct ways which lead to different sets of classification results. It is also expected that there should exist a set of suitably chosen transformations, with the corresponding re-shaped histograms attaining better classification results then the original ones. Specifically, we define a new random variable z, which takes values in the set $Z = \{z_1, \ldots, z_r, \ldots, z_R\}$, which is the result of applying a non-linear transformation $T(\cdot)$ to x, i.e., $z = T(x)$. The resulting pmf $P(z_r) = \Pr\text{ob}(z = z_r)$ of the transformed variable z is given by $P(z_r) = \text{Prob}(z = z_r) = \sum_{u: z_r = T(x_u)} P(x_u)$.
From this equation, it is seen that, if we consider the histogram as an approximation of the original pmf, the transformation T can be indirectly realized by accumulation of selective bin counts into a single bin. As a result, the simplicity and efficiency of the original compressed domain approach can be retained.
In addition to this transformation, we can further specify a suitable subset of the random variable domain to further improve the content characterization capability of the transformed random variable. In other words, we need to construct a *conditional* pmf as the classifier input.
Our objective is thus to select multiple intervals $I_1, \ldots, I_m, \ldots, I_M$ from the original variable domain such that $I_{m_1} \cap I_{m_2} = \phi$ where $m_1 \neq m_2$. Given these selected intervals $I = \bigcup_m I_m \subset Z$, the bin counts within the interval would be accumulated together to form a single count. For $z_r \in I$, the conditional pmf (cpmf) is defined as $P(z_r \mid I) = \frac{P(z_r)}{\sum_{z_r \in I} P(z_r)}$ which satisfies $\sum_r P(z_r \mid I) = 1$.
In other words, our objective is to select the optimal transformation and the associated interval subsets to construct a cpmf as classifier input, such that the classification accuracy is maximally improved. However, it can be expected that there are a large number of possible ways to construct the cpmf. To effectively explore the

search space, we adopt evolution strategy (ES) as our search technique.

4. IMAGE CLASSIFICATION

As mentioned previously, due to the large size of current image databases, effective classification of images into different groups is essential for content-based image retrieval applications to avoid full traversal of the database for a specific query. Suppose there are K classes of images labeled as $G_1, G_2, \ldots, G_K$ in the database. Each image e_n, $n=1,\ldots N$ in the database is assigned a class membership label $k \in \{1,\ldots\ldots,K\}$ which indicates the class it belongs to. With the correct image class structure as ground truth, we try to approximate this structure using only compressed domain features. As expected, it is found that there is a discrepancy between classifications based on compressed domain features and the original image class structure. Due to this discrepancy, we propose to transform the compressed domain features such that this discrepancy is minimized.

5. EVOLUTION STRATEGIES BASED OPTIMIZATION

5.1. Evolution Strategy

Evolution Strategy (ES) [5] is one of the most commonly used optimization algorithms among different evolutionary algorithms. The fundamental principle of the algorithm is to search for an optimal solution in a population of candidates based on a fitness function. In general, evolutionary algorithms are most effective for non-linear optimization problems with a large search space. Coincidently, our problem, which is to search for an optimal transformation of the compressed domain features, is also non-linear with a large number of potential solutions, and this motivates our choice of ES.

5.2. Representation of candidate transformations

Each candidate transformation in a population is represented by a real-valued vector $w^s, s=1,\ldots\ldots,S$, where S is the size of the population. The length of the vector is equal to half of the total number of bins L of the original histogram. Each component of the vector is bounded within the interval $[0,1]$ and a specific transformation is realized by observing adjacent pairs of components in the vector. Specifically, the components in the vector are sorted in ascending order where each pair of adjacent values in the sorted string is used to define a selected interval. For each pair of adjacent components (h_1^s, h_2^s), a corresponding interval bounded by these two values will be selected. The bin counts within the interval will be merged together to form a single count. Then all these bin counts will be concatenated and renormalized to form the cpmf. Figure 1 shows an example to illustrate the transformation process.

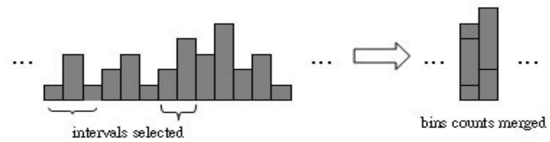

Figure 1, Illustration of the transformation process.

5.3. Fitness Function

Since a global optimal transformation may not exist for all the classes, we adopt a multiple classifier approach where each classifier is associated with a *locally* optimal transformation for a specific class. As a result, we can individually search for the optimal local transformations for each class using ES by partitioning the original population into sub-populations associated with each class. The main requirement for the k-th classifier and its associated transformation is that it should be able to accurately identify images of class k, while at the same time able to avoid misclassifying images from other classes as those of class k.

For each class G_K and a specific candidate transformation $T^{k,s}$ in its associated sub-population, we can define the following prototype conditional pmf $\overline{P}^k(z_r \mid I^{k,s}) = \frac{1}{|G_k|}\sum_{n:e_n \in G_k} P_n(z_r \mid I^{k,s})$ where $|g_k|$ is the cardinality of class g_k, and $I^{k,s}$ is the selected interval subset of the candidate transformation. A new class structure can be imposed on the images based on the prototype cpmfs as follows:

$$e_n \in G_k^s \text{ if}$$
$$d(P_n(z_r \mid I^{k,s}), \overline{P}^k(z_r \mid I^{k,s})) \le d(P_n(z_r \mid I^{k,s}), \overline{P}^l(z_r \mid I^{k,s}))$$
$$where\; l = 1,\ldots,K, l \ne k$$

$d(\cdot,\cdot)$ is the euclidean metric.

To measure the conformance of this imposed class structure based on compressed domain features with the original class structure, we define the fitness function as $\Phi_k^s = (|G_k \cap G_k^s| + |G_k' \cap G_k^{s'}|)$ where G_k' and $G_k^{s'}$ denote the complements of G_k and G_k^s respectively. In other words, the first term counts the correct positive classification and the second term counts the correct negative classification.

6. EXPERIMENTAL RESULT

To evaluate our proposed approach, we performed simulation experiments using a database of 1650 images which are classified into 8 classes based on their visual contents. Figure 2 shows the representative images from the classes. From the database, half of the images are randomly selected and used for system training while the remaining half are used for testing the current approach. We then apply ES to search for the optimal transformations for both the color and edge orientation histograms.

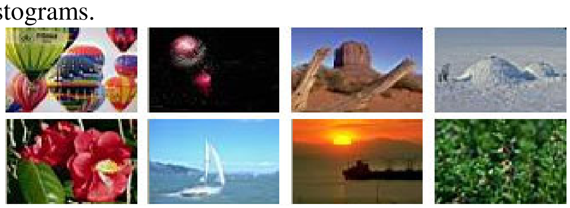

Figure 2, representative images from different classes.

For system training, a $(\mu+\lambda)$ES strategy is applied where μ parents are used to generate λ offspring through mutation and recombination. The elitist selection mechanism is adopted where both parents and their offspring will compete to survive into the next generation. The population size was set to 20 and the maximum number of generation was set to 500.

Generally, for ES, mutation is the dominant search operator while recombination serves a supporting role. In our case, the isotropic form of mutation is implemented and global recombination is used to generate new candidates where all μ individuals are allowed to take part in the process. The algorithm will be terminated if there is no improvement of the fitness score over a prescribed number of generations.

As there are *k* classifiers with their possibly conflicting outputs, we need to implement a fusion mechanism so that only one final class label will be assigned to each image. Specifically, the *k*-th classifier can produce an ascending list of distance measure values $d_n^{k,l}, l = 1,\ldots,K$ of the histogram to each class prototype according to the following equation:

$$d_n^{k,l} = d(P_n(z_r \mid I^k), \overline{P}^l(z_r \mid I^k))$$

With the ordered list, we can define a classifier score for each classifier $\alpha_n^k = \frac{\sum_{l \neq k} d_n^{k,l}}{d_n^{k,k}}$. A high score will be obtained if the image histogram is close to the center associated with the *k*-th classifier and far away from other centers. The final class membership $\gamma(e_n)$ of the image e_n will be determined as follows: $\gamma(e_n) = \arg\max_k \alpha_n^k$

Table 1 shows the classification result on the database based on our proposed approach. It can be seen that both the classification accuracies of the training and test sets have been significantly improved as a result of the evolutionary optimization process.

Class	Training Set		Test Set	
	Before Transf.	After Transf.	Before Transf.	After Transf.
1	87.23%	87.23%	85.11%	87.23%
2	89.02%	93.29%	93.87%	98.16%
3	84.91%	88.68%	94.34%	94.34%
4	92.22%	92.81%	92.77%	95.78%
5	88.41%	90.58%	90.58%	91.30%
6	93.33%	96.67%	96.67%	97.78%
7	92.41%	96.20%	93.67%	96.20%
8	89.66%	89.66%	96.30%	96.30%
Overall	89.76%	92.32%	92.56%	94.51%

Table 1, Classification results.

Please note that the ES is a one-off process which once the optimized transformation is found, it can be applied to the whole image database without requiring further optimization. In addition, the transformation only requires addition and removal of bin counts in the original histogram, thus no expensive computation is required.

7. CONCLUSION

In this paper, we propose a simple transformation mechanism for compressed domain features such that the content characterization capability of these features can be improved without compromising the efficiency of these approaches. To effectively search for an optimal transformation from a large solution space, we adopt Evolutionary Strategy as the search technique. Experiments have shown that our proposed approach is able to enhance the content characterization capability of the compressed domain features, as seen from the improved classification rates for the various image categories. At the same time, the simple operations of histogram bin count merging for implementing the transformation also ensures that the original efficiency of these approaches is retained.

8. REFERENCES

[1] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Compression Standard*. New York, NY: Van Nostrand Reinhold, 1993.

[2] Guocan Feng, Jianmin Jiang, “JPEG compressed image retrieval via statistical features”, *Pattern Recognition*, vol. 36, pp. 977-985, 2002

[3] Yu-Len Huang and Ruey-Feng Chang, “Texture features for DCT-coded images retrieval and classification”, *Proc. IEEE Int. Conf. on Acoustics, Speech & Signal Processing,* vol. 6, pp. 3013-3016

[4] Shen, B. and Sethi, I.K., “Direct feature extraction from compressed image,” SPIE, Vol.2670 *Storage & Retrieval for Image and Video Database IV*, 1996

[5] Hans-Georg Beyer, “The Theory of Evolution Strategies”, Springer, 2001.