TRANSFORM FEATURES FOR TEXTURE CLASSIFICATION AND DISCRIMINATION IN LARGE IMAGE DATABASES

John R. Smith and Shih-Fu Chang

Center for Telecommunications Research and Electrical Engineering Department
Columbia University, New York, NY 10027, U.S.A.

Abstract

This paper proposes a method for classification and discrimination of textures based on the energies of image subbands. We show that even with this relatively simple feature set, effective texture discrimination can be achieved. In this paper, subband-energy feature sets extracted from the following typical image decompositions are compared: wavelet subband, uniform subband, discrete cosine transform (DCT), and spatial partitioning. We report that over 90% correct classification was attained using the feature set in classifying the full Brodatz [3] collection of 112 textures. Furthermore, the subband energy-based feature set can be readily applied to a system for indexing images by texture content in image databases, since the features can be extracted directly from spatial-frequency decomposed image data.

In this paper, we also show that to construct a suitable space for discrimination, Fisher Discrimination Analysis [5] can be used to compact the original features into a set of uncorrelated linear discriminant functions. This procedure makes it easier to perform texture-based searches in a database by reducing the dimensionality of the discriminant space. We also examine the effects of varying training class size, the number of training classes, the dimension of the discriminant space and number of energy measures used for classification. We hope that the excellent performance for texture discrimination of these simple energy-based features will allow images in a database to be efficiently and effectively indexed by contents of their textured regions.

1. INTRODUCTION

With the increased prevalence of digital image and video archives, new techniques are being investigated to perform efficient searching and retrieval of visual data. It is being realized that traditional text-based indexing schemes alone are not sufficient when the data is visual. Upon translation of visual features into textual description, problems of consistency and completeness arise. Since information exists at many semantic levels within a visual scene, it is impossible to describe all levels sufficiently. Furthermore, text-based schemes require human assistance in describing textually the pertinent information contained in visual scenes. While this would require tremendous human effort for large image collections, the sheer number of unique frames or scenes in a video sequences makes any human-assisted indexing scheme impractical. Consequently, there has been a new focus on automated visual content-based approaches towards indexing images and video.

1.1 Content-Based Visual Query

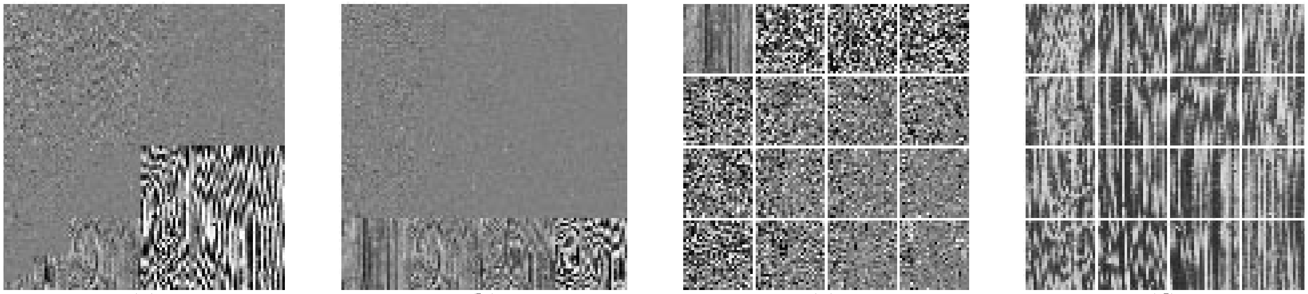
With content-based techniques, the important visual features of image and video data are described mathematically using feature sets that are derived from the digital data. If chosen properly, the feature sets may coincide well with intuitive human notions of visual content while providing effective mathematical discrimination. Furthermore, the features may be extracted automatically by computer without requiring human assistance. This approach allows the database to be indexed using the discriminating features that characterize visual content and searched using visual keys. A content-based search of the database proceeds by finding the items most mathematically and visually similar to the search key.

Characterizations of *texture*, color and shape using feature sets are beginning to find application towards content-based approaches for large image and video databases [8][9]. Other features such as object spatial relationship and object tracking and motion in video also seem promising in content-based systems. This paper focuses solely on the efficient characterization of *texture* as a first step towards a visual content-based query system.

The feature sets for texture examined here are computed directly from image spatial-frequency blocks which can be obtained from several popular image decompositions. When these feature sets are combined with image segmentation, each uniquely textured region of the images in the database can be characterized and added to a texture-based index for the database. We hope that the feature sets derived here can be generalized to universal unconstrained image and video database applications.

1.2 Texture Classification Experiments

In the experiments reported in this paper, each of the 112 Brodatz [3] texture images was randomly cut into rectangu-



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

FIGURE 1. Image decompositions, (a) Wavelet, (b) Uniform subband, (c) Mandala DCT and (d) Spatial block.

lar pieces of random sizes. A fraction of the texture cuts were used for training to produce Fisher discriminant functions used for classifying the remaining texture cuts. Using this procedure, the classification performance of the energy-based feature sets was measured using subbands of several image spatial-frequency decompositions. Classification rates of over 90% were obtained for both wavelet subband and uniform subband image decompositions. DCT reported just over 80% classification rate, while simple spatial block features gave only 34% correct classification.

2. IMAGE TRANSFORM FEATURES

The following image decompositions were used to produce the energy-based features sets, see Figure 1:

- wavelet subband (5 iterations, 16 subbands),
- uniform subband (4x4, 16 subbands),
- DCT (4x4, terms sorted to produce 16 subbands),
- spatial partition (4x4, 16 blocks).

Each of the decompositions produced 16 subbands or image blocks. For each decomposition the energies were measured by calculating the variance and mean absolute value of each subband. This produced a feature vector of 32 terms to describe each texture image. For each type of image decomposition, the 32 term feature vectors generated by the training data were mapped into a reduced space using Fisher Discriminant Analysis. In addition to compacting the feature space, this also sets up the mechanism by which the remaining test texture cuts are classified. By measuring the success rate of classification, the efficacy of the original features in providing texture discrimination can be determined.

3. FISHER DISCRIMINANT ANALYSIS

Fisher Discriminant Analysis generates a family of linear composites from the original feature vectors that provide for maximum average separation among training classes. Although the resultant composites are not orthogonal, they are uncorrelated with each other [5]. Fisher Discriminant Analysis works by finding the eigenvectors of scatter matrices which describe class separability. The criteria of class separability is formulated using the *within-class* (W), *between-class* (B), and *total* (T) scatter matrices such that $W^{-1}B$ represents the ratio of *between-class* to *within-class* sum-of-squares for K classes. This matrix is then used in place of traditional feature covariance matrix through the procedure of Principal Component Analysis [5].

The classification is performed by using the subset of the resulting eigenvectors, or discriminant functions, that account for the largest total variation. A minimum distance rule, such as the square distance in the discriminant space, is used to assign new observations to the nearest classes, see Figure 2. Overall, to classify an unknown texture cut, it is first decomposed using the appropriate filter bank or image transformation. Next, the subband energies are measured to produce a feature vector that describes the texture. Then the Fisher discriminant transformation maps the feature vector to the trained discriminant space. Finally, the distance function is used to assign the texture cut to the appropriate class.

4. CLASSIFICATION RESULTS AND DISCUSSION

From each of the 112 Brodatz texture images, 20 randomly sized and positioned rectangular cuts were made to produce 2240 total texture cuts. For training purposes, from 1 to 10 cuts from each class were used in the experiments. The remaining 10 cuts from each class which were set aside for later classification, comprised the test set.

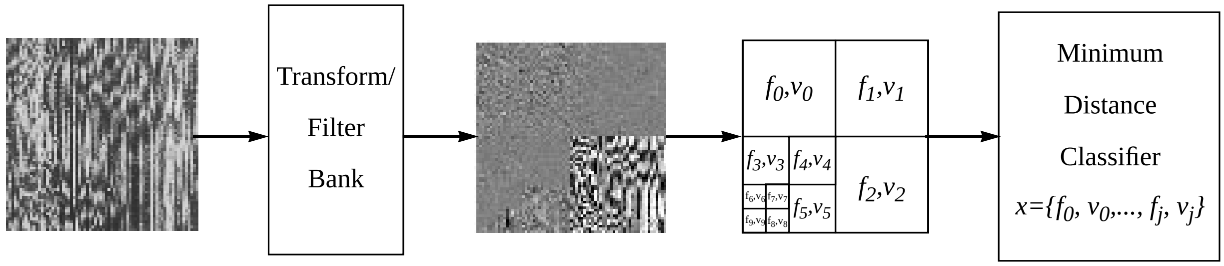


FIGURE 2. Texture classifier for Brodatz textures cuts using subband-energy based features.

The collection of Brodatz textures consists of textures of both statistical and structural natures. Structural textures are considered to consist of texture primitives which are repeated systematically within the texture. In statistical textures usually no repetitive structure can be identified. Typically, to describe textures mathematically, different techniques have been applied separately to characterize structural and statistical textures. The spatial-frequency subband energy approach used in the paper was applied to both types of textures.

The cuts made from the Brodatz texture images to produce the training and testing sets were much smaller than the original images. Since the cuts grabbed only part of the original textures, the cuts did not always capture the original textures well. In a database of real images, there may be similar difficulty in obtaining homogeneous representatives of textured regions. Even though this has a negative impact on the texture classification, we did not eliminate any Brodatz textures from the classification experiments.

4.1 Comparison of Transform Feature Sets

The classification performance of the energy-based feature sets is summarized in Table 1. Here, five training cuts from each texture class were used for training. Testing was performed using the discriminant functions based on the subband-energy feature set to classify the remaining ten cuts from each class.

Notice that energy-based features derived from wavelet subband and uniform subband decompositions performed equally well. This is interesting considering that they partition the frequency plane differently. The similarity in performance is due to the fact that textures contain a significant amount of energy in the middle frequencies in addition to low frequencies. By iterating on the lowest frequencies, the wavelet decomposition does not capture the mid-frequency range well. Likewise, by splitting the frequency spectrum into uniformly spaced bands, the uniform decomposition resolves the middle frequencies better than wavelet decomposition, but does not capture low frequencies well. In our experiments for textures, the trade-offs are comparable.

A slight reduction in performance was seen with the Haar two-tap filter wavelet decomposition compared to the QMF16c [7] wavelet. This is a result of the inferior stopband characteristics for the Haar filter. The DCT decomposition produced even further reduction in performance. Since the DCT originates from spatial blocks, the resulting frequency spectrum suffers from inter-band energy leakage [1], which produces a poorer feature set for texture discrimination. The spatial block case used for comparison shows the worst performance. This makes apparent that texture information cannot be captured well from simple spatial block measures.

Wavelet QMF16c	Wavelet Haar	Uniform Subband QMF16c	4x4 DCT Mandala block	4x4 Spatial
92.14%	90.17%	92.14%	85.18%	34.65%

TABLE 1. Classification performance of various transform feature sets.

4.2 Effect of Training Class Size

As should be expected, the classification rates increase as the number of texture cuts from each class used for training increases. This is indicated in Figure 3 for the different image decompositions. As more training cuts are used per class, each class is more accurately defined. And as the sample statistics approach the true underlying distributions for each class, the classification performance increases. In these experiments good performance was reached using relatively few training samples (2 to 4) per class and the subband-energy feature set.

4.3 Effect of Number of Training Classes

In an image database application, all possible texture classes cannot be available to generate the discriminant functions. Therefore, one would like to find discriminant functions from a subset of texture classes that are useful in general for most textures. To determine how well the trained discriminant functions might extend to other textures, a random subset of Brodatz texture classes was used for training. Then the discriminant functions were used to classify cuts from all 112 Brodatz texture classes. The results are shown in Figure 4. Notice that even when only a quarter of the number of classes were used for training, performance shows only slight degradation. This indicates that the subband-energy feature might generalize well to textures outside the Brodatz collection.

4.4 Effect of Number of Discriminant Functions

The Fisher discriminant functions that account for the largest separation are indicated by having the largest normalized eigenvalues [5]. By using a subset of the discriminant functions that still significantly separate the training classes, the dimension of the feature space can be reduced. This reduces the computation needed to measure texture similarity, and also reduces the dimension of the search space in a database application. It is advantageous to find feature sets which allow for the highest energy packing possible for classification. Figure 5 shows the results of classification, using most significant subsets of the discriminant functions for classification. Notice that performance dropped only slightly, even when using just a quarter of the discriminant functions.

4.5 Effect of Number of Features

When correlated features are added to the feature set, Fisher Discriminant Analysis removes redundancy and produces a set of uncorrelated discriminant functions. In general, any measures that provide some degree of class separation should be included in the feature set [6]. However, as more features are added, there is a trade-off between classification performance and computation.

In these experiments, both mean absolute value and variance of frequency bands were used as energy measures in the feature sets. However, these features show some correlation. In fact, when using zero mean filters, the high frequency subbands have zero mean. In this case, the variance measure is actually just the sum-of-squares, which makes it even more similar to the sum of magnitudes measure. Never-the-less, when both energy measures were included in the feature set here, the overall classification rate increased over the performance using just one energy measure, see Figure 6.

5. CONCLUSION

In this paper, texture classification was performed using subband energy-based feature sets from several image decompositions. We used the entire set of 112 texture classes and were able to achieve over 90% correct classification. We showed that classification performance depends on the underlying decomposition used to obtain the spatial-frequency blocks. We also showed the dependency on the number of training cuts used per class, number of discriminant functions and number of features used. We also reported classification results using a subset of texture classes for training, which more closely matches the environment of a database of real images. In future work, we will extend these results to real images for the purpose of performing texture segmentation and indexing in image and video database applications.

References:

[1] Ali N. Akansu and Richard A. Haddad, *Multiresolution Signal Decomposition -- Transforms, Subbands and Wavelets*, Harcourt Brace Javanovich, 1992.

[2] R. Barber, W. Equitz, M. Flickner, W. Niblack, D. Petrovic, P. Yanker, “Efficient Query by Image Content for Very Large Image Databases,” *COMPCON*, San Francisco, 1993.

[3] P. Brodatz, Textures: a Photographic Album for Artists and Designers, Dover, New York, 1965.

[4] N. Dal Degan, R. Lancini, P. Migliorati and S. Pozzi, “Still images retrieval from a remote database: The system *Imagine*,” *Signal Processing: Image Communications*, vol. 5, 1993, pp 219 - 234.

[5] William R. Dillon and Matthew Goldstein, *Multivariate Analysis*, John Wiley & Sons, 1984.

[6] Keinosuke Fukunaga, *Introduction to Statistical Pattern Recognition*, Harcourt Brace Javanovich, 1990.

[7] J.D. Johnston, “A Filter Family for Use in Quadrature Mirror FIlter Banks,” *Proc. ICASSP*, pp. 291-294, 1980.

[8] Wayne Niblack, “Querying large image databases by content,” *OE Reports of The International Society for Optical Engineering*, No. 123, March 1994, pp. 15.

[9] Wayne Niblack, et. al., “The QBIC Project: Querying Images by Content using Color, Texture, and Shape,” IBM RJ 9203 (81511), February 1, 1993.

[10] R. W. Picard, T. Kabir, and F. Liu, “Real-time Recognition with the entire Brodatz Texture Database,” *M.I.T. Media Laboratory Perceptual Computing Group Technical Report*, June 1993, No. 215.

[11] T. R. Reed and J.M. Hans du Buf, “A Review of Texture Segmentation and Feature Extraction Techniques,” *Computer Vision, Graphics, and Image Processing*, 1993, Vol. 57, pp. 359--372.

[12] John R. Smith and Shih-Fu Chang, “Quad-Tree Segmentation for Texture-Based Image Query,” accepted for ACM Multimedia, October, 1994, San Francisco, CA.

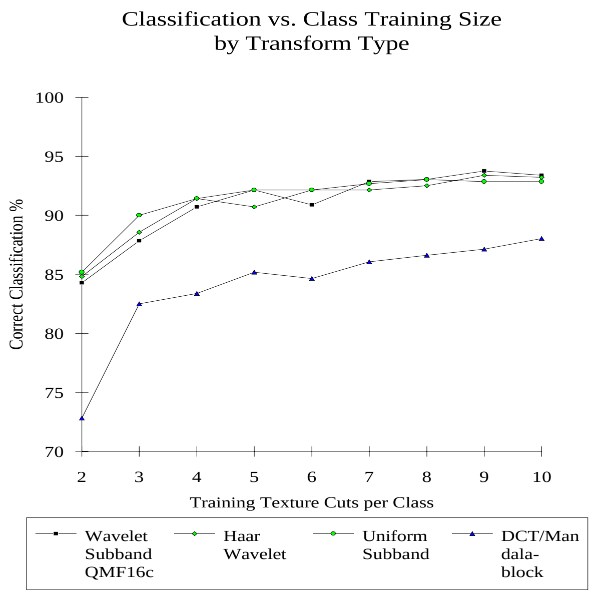


FIGURE 3. Effect of training class size on classification rate for 112 Brodatz texture classes.

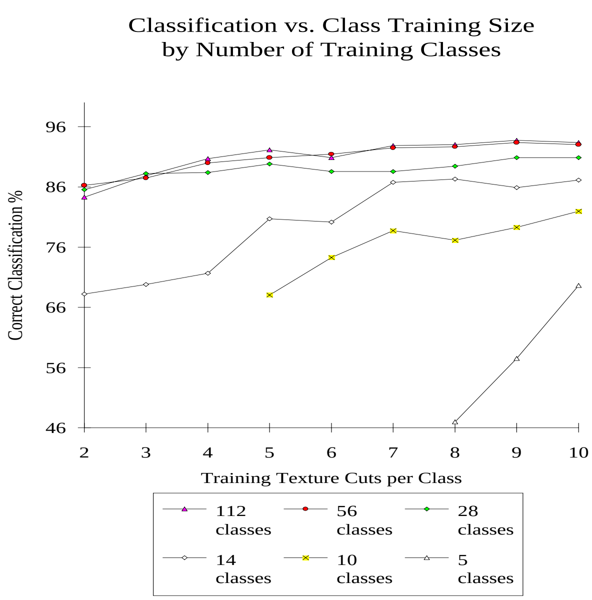


FIGURE 4. Effect of number of training classes on classification rate for 112 Brodatz texture classes (QMF16c Wavelet decomp).

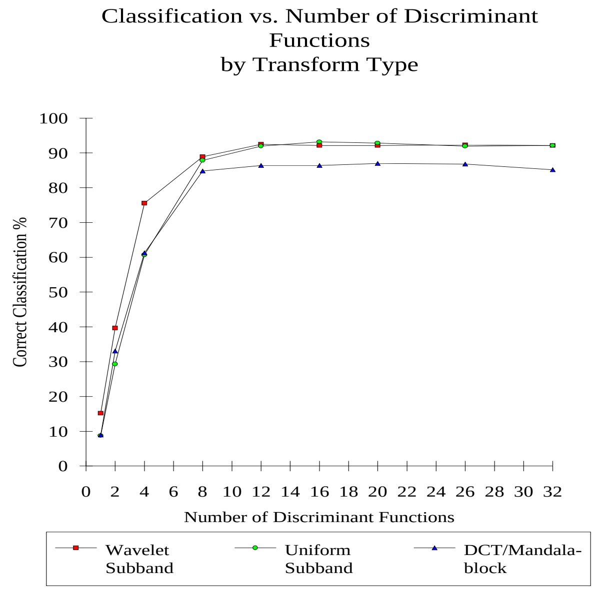


FIGURE 5. Effect of the number of discriminant functions used for classification on classification rate for 112 Brodatz texture classes.

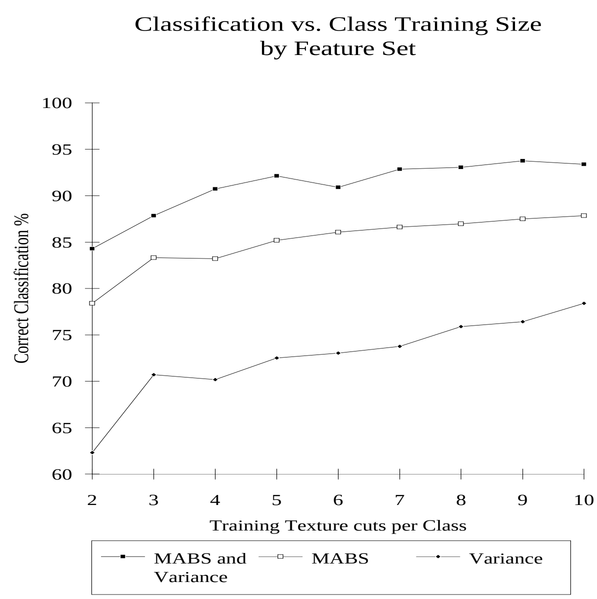


FIGURE 6. Effect of feature set size on classification rate for 112 Brodatz texture classes (QMF16c Wavelet decomp).