AUTOMATED BINARY TEXTURE FEATURE SETS FOR IMAGE RETRIEVAL

John R. Smith and Shih-Fu Chang

Columbia University Department of Electrical Engineering and
Center for Telecommunications Research
New York, N.Y. 10027

ABSTRACT

Digital image and video libraries require new algorithms for the automated extraction and indexing of salient image features. Texture features provide one important cue for the visual perception and discrimination of image content. In this paper we propose a new approach for automated content extraction that allows for efficient database searching using texture features. The algorithm automatically extracts texture regions from image spatial-frequency data which are represented by binary texture feature vectors. We demonstrate that the binary texture features provide excellent performance in image query response time while providing highly effective texture discriminability, accuracy in spatial localization and capability for extraction from compressed data representations. We present the binary texture feature extraction and indexing technique and examine searching by texture on a database of 500 images.

1. INTRODUCTION

In this paper we propose and evaluate an algorithm for the automated extraction of texture information which enables efficient filtering of and searching through large collections of digital images and videos using texture features. The texture features are produced by thresholding and morphologically filtering image spatial/spatial-frequency (*s/s-f*) subbands. Texture is represented by a binary feature set whereby each element in the binary set indicates the energy relative to threshold in a corresponding *s/s-f* subband. More specifically, each *s/s-f* channel is represented by a unit length binary feature vector b_k. An orthogonal *s-f* decomposition is used which allocates the *s-f* information such that the subbands are independent. As such, $\{b_k\}$ forms a set of linearly independent elementary texture vectors which provides a basis for the binary texture space. Texture feature vectors are composed of binary combinations of vectors from the texture basis $\{b_k\}$.

The process of texture extraction from *s/s-f* subbands by thresholding and morphological filtering identifies spatially localized and arbitrarily shaped regions of texture within each image. The spatial localization property preserves the logical spatial relationships between texture regions. This provides for a secondary criterion for evaluating image matches. Since the extraction algorithm preserves texture region shape, region boundary descriptions may also be used in the match criteria. Furthermore, since texture is represented by binary feature vectors, the texture regions may be indexed very efficiently which helps to minimize image query response time. Finally, the algorithm allows for texture extraction from compressed orthogonal *s/s-f* representations of image and video data [3][4]. Feature extraction in the compressed-domain provides great potential for reducing computational complexity because it avoids the costly operation of decompressing the images and videos. This is particularly relevant for large image and video databases which may have hundreds of thousands of image and video items stored in compressed form.

2. TEXTURE FEATURES

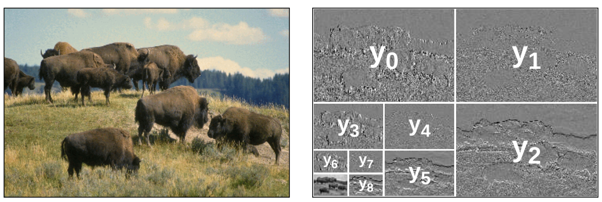
Texture refers to a visual pattern that has properties of homogeneity that do not result from the presence of only a single color or intensity. Previous attempts at modeling texture include the following approaches: random field modeling [11], co-occurrence matrices [8] and *s-f* techniques [5][10] [13] which include in particular, Gabor filters [2][9][12]. Thus far, no single best texture model has been identified. However, there is evidence that early human vision uses receptive field units tuned to orientations and *s-f*s [7]. In particular, models of the human visual system that use Gabor filters to model the receptive fields sufficiently account for psychophysical data obtained in texture discrimination experiments [6].

2.1. Gabor functions

Gabor filters produce *s-f* decompositions that achieve the theoretical lower bound of the uncertainty principle. They attain maximum joint resolution in space and *s-f* bounded by the relations $\Delta_x^2 \cdot \Delta_u^2 \geq \frac{1}{4\pi}$ and $\Delta_y^2 \cdot \Delta_v^2 \geq \frac{1}{4\pi}$ where $[\Delta_x^2, \Delta_y^2]$ gives resolution in space and $[\Delta_u^2, \Delta_v^2]$ gives resolution in *s-f*. This is highly significant in the process of texture extraction in which the conflicting objectives of accuracy in texture representation and texture spatial localization are both important. In addition to good performance in texture discrimination and segmentation [2], the justification for Gabor filters is also supported through psychophysical experiments. Beck [1] demonstrated that human texture segregation results from information corresponding to outputs of *s-f* channels. Texture analyzers implemented using 2-D Gabor functions produce a strong correlation with actual human segmentation [12]. Furthermore, the receptive visual field profiles are adequately modeled by 2-D Gabor filters [6]. However, Gabor functions are both nonorthogonal and complex which makes implementation very difficult.

2.2. Wavelet subband features

A more practical way to gain in the trade-off between space and *s-f* resolution without using Gabor functions is with a dyadic or wavelet filter bank. The wavelet filter bank produces octave bandwidth segmentations in *s-f*. It allows simultaneously for high spatial resolution at high *s-f*s and high *s-f* resolution at low *s-f*s. Furthermore, the wavelet tiling is supported by evidence that visual *s-f* receptors are spaced at octave distances [6]. A two-band quadrature mirror (QMF) filter bank utilizes orthogonal analysis filters to decompose data into low-pass and high-pass frequency bands. When filtering is recursively applied to the low-pass frequency bands the QMF filter bank produces a octave band split or wavelet decomposition as illustrated in Figure 1. Separable QMF filters reduce the computational complexity of the filter banks and make them very attractive for implementation.



(a) (b)

Figure 1. Extraction of texture features, (a) "Buffaloes" image, (b) QMF wavelet transformed image, filter outputs, y_k, correspond to *s-f* subbands.

A QMF wavelet filter bank was used for texture classification by Kundu and Chen [10]. The authors identified several aspects of the QMF filter bank as being relevant to texture extraction: orthogonality and completeness of the basis functions, filter outputs that are spatially localized and the reduction of complexity afforded by decimation of filter outputs. In the classification of Brodatz textures the QMF wavelet features performed better than those proposed by Haralick [8]. Furthermore, in [13] we evaluated the performance of texture features sets derived from several *s-f* decompositions including QMF wavelet and uniform subband and DCT. The texture feature sets were evaluated on the basis of their ability to classify texture cuts made from all 112 Brodatz textures. We found that a 9-dimensional feature set derived from the *s-f* energies provided for over 90% successful classification.

3. AUTOMATED TEXTURE ANALYSIS

The automated segmentation of texture images and extraction of texture regions is essential for content-based image retrieval. Texture segmentation is the process by which the image is split into different regions of homogeneous texture. By using a binary feature set derived from image *s-f* subbands, an effective algorithm for texture segmentation is obtained. The procedures for texture representation and texture extraction are described next.

3.1. Texture feature set notation

The texture features are represented as follows: Let $x[m,n]$ be the discrete 2-D representation of the image. Let $h_k[m,n]$, where $k = 0 \dots N-1$ be the set of 2-D orthogonal filters that compose the wavelet filter bank. The 2-D convolution,

$$y_k[m,n] = x[m,n] * h_k[m,n] \quad (1)$$

generates the output of subband k by approximately band-limiting $x[m,n]$ to an exclusive range of *s-f*s. Let τ_k be the threshold in subband k such that,

$$b_k[m,n] = \begin{cases} 1 & \text{if } |y_k[m,n]| \geq \tau_k \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Definition 1 *Texture Feature Set. Let $\mathcal{B}^N$ be the N dimensional binary space spanned by elementary binary feature vectors b_k, each of which corresponds to the thresholded output of the respective s-f filter k. A texture feature set $\hat{t}$ is defined as a vector in binary space $\mathcal{B}^N$ and is composed of binary combinations of basis vectors from $\{b_k\}$.*

The objective of the texture extraction technique is to utilize the *s-f* decompositions provided by the filters h_k to produce a characterization of texture. The design parameters include the filters h_k and the energy thresholds τ_k.

The 2-D matrix of texture vectors obtained by concatenating the thresholded subbands denotes the texture value at each point from the image $x[m,n]$. This pixel-level description of texture is not immediately useful for image retrieval because texture is typically perceived as a regional process. As such, the texture points are grouped into regions using non-linear filtering operations as will be explained below. The obtained texture regions are represented by binary texture vectors $\hat{t}$. Then searching for images using texture content is accomplished by referring to the texture vector values corresponding to regions within the images.

3.2. Texture feature extraction

As illustrated in Figure 2, the image $x[m,n]$ is fed into the size N filter bank. The filter outputs are maximally decimated and energy magnitudes are computed. This obtains the band limited *s-f* energies within dyadic neighborhoods around each image pixel. By computing fixed-sized neighborhood energies on the decimated images, the actual image pixel neighborhoods are sized in accordance with the uncertainty principle to obtain maximum statistical accuracy in the measurement of *s-f* energy. The energy images are illustrated for the "Barbara" image in Figure 4(b).

Next, for all channels, $k = 0 \dots N-1$, the output of filter k is thresholded at level τ_k to produce bi-level image $b_k[m,n]$. Each bi-level image has a non-zero value at points of high energy that correspond to lines, edges, noise and textures within the original image. By morphologically filtering each bi-level image $b_k[m,n]$ to produce $f_k[m,n]$, the edges, lines and noise are reduced while the texture content is enhanced. This is accomplished through median filtering followed by sequential labeling and spatial size thresholding. The outputs resulting from these operations are illustrated in Figure 4(c).

Since the morphological filtering is performed on bi-level images, the median statistic is equivalent to the mode. By

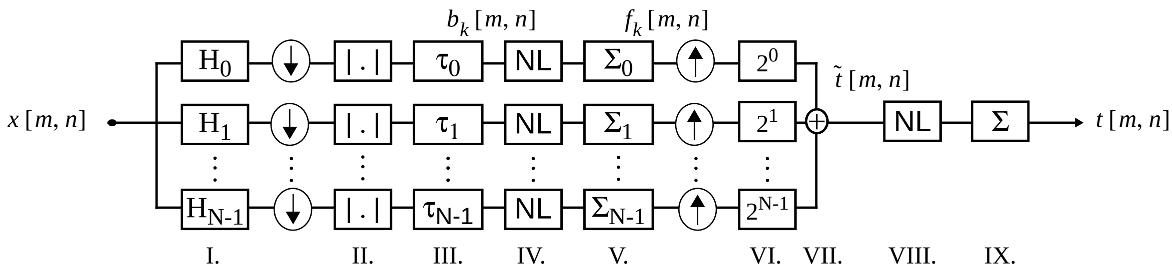


Figure 2. Extraction of textures using wavelet filter bank, (I.) H_k are orthogonal filters, (II.) $|\cdot|$ obtains s-f energy, (III.) τ_k are energy thresholds, (IV.) non-linear filtering (NL), (V.) spatial size thresholds (Σ_k), (VI.) expansion into binary space $\mathcal{B}^N$, (VII.) concatenation of channels, (VIII.) final NL filtering and (IX.) size thresholding (Σ).

choosing a box shape for the filter, the points corresponding to the lines, edges and noise are eliminated because they do not fill the majority of the window. However, since filter outputs are s-f band-limited and maximally decimated before thresholding, a window majority indicates that the center point belongs to a region with a homogeneous s-f texture characteristic.

From the morphologically filtered bi-level images $f_k[m,n]$ the image $\tilde{t}[m,n]$ is formed by upsampling using pixel replication, and concatenation of the $f_k[m,n]$'s. This operation composites the elementary texture regions from the bi-level subband images. Equivalently, this composes texture feature vectors as binary combinations of the elementary binary texture basis functions. Next, median filtering is performed on $\tilde{t}[m,n]$ to allow dominant texture regions to be enhanced. A region size threshold is also applied to eliminate the remaining small regions. This produces the binary texture vector image $t[m,n]$ as illustrated in Figure 4(d). This image contains the salient regions of texture found from the original image. Each region is represented by a binary texture vector that indicates the significant s-f content in each region. The regions corresponding to extracted textures are illustrated in Figure 4(e).

A pedagogical example of texture extraction on a simple image using a three channel filter bank is illustrated in Figure 3. The outputs of the filter bank channels after thresholding and nonlinear processing, as described above, are illustrated in Figure 3(a). Each channel corresponds to a set of s-fs and a unique direction in the 3-D binary feature space. Binary combinations of the channel outputs produce $2^3 = 8$ texture extraction images as illustrated in Figure 3(b).

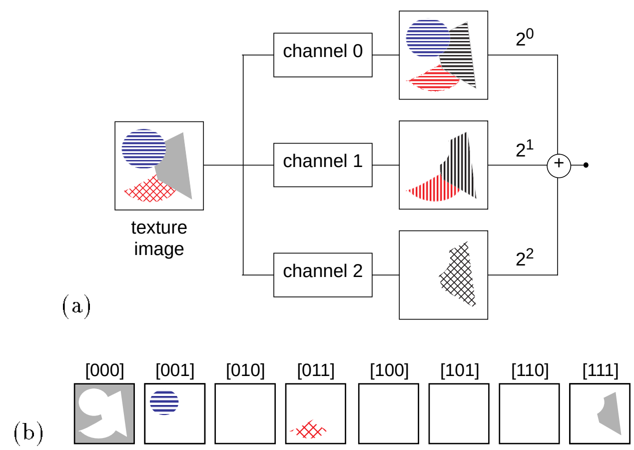


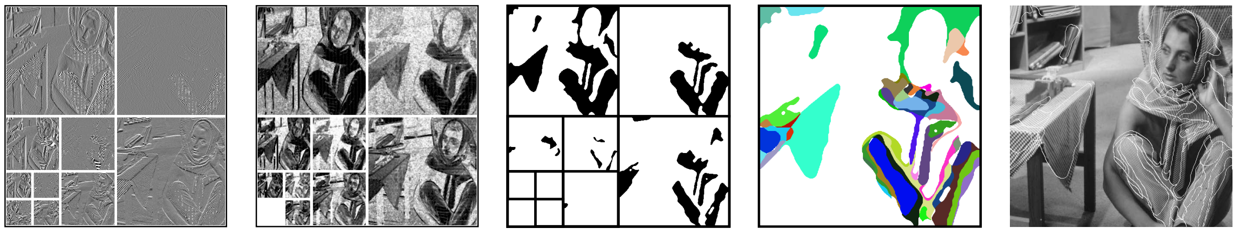
Figure 3. Example of texture extraction using three channel filter bank, (a) the thresholded and processed images for the three channels are concatenated, which produces (b) eight binary texture images that illustrate the segmented texture regions.

3.3. Texture retrieval

The texture feature set effectiveness was evaluated using examples of image retrieval based on texture content. The performance of the s-f texture features was previously evaluated using the Brodatz texture classes in [13] and they were found to achieve accurate texture classification. Here, we evaluate the discriminability and texture matching capacity afforded by the texture features sets by composing queries of the image database using texture.

We allow that texture query feature sets can be formed in two ways. In the first, a sample image texture is provided from which the query texture vector is computed directly. Alternatively, each of the s-f subbands is assigned a visual pattern that illustrates its scale and orientation. A query texture is composed by selecting from the depicted texture subband elements. In both cases, the search process is accomplished by accessing indexes defined over the binary texture space. In the N-dimensional binary texture space, texture regions are indexed by using a depth N binary tree. The efficient indexing of the binary texture feature vectors provides considerable advantage over other texture matching procedures. These other real-valued texture feature vectors typically require a distance computation to evaluate texture closeness. This is expensive, as it may be necessary to compute the distance to each item in the database in response to a query.

The results of a texture-based query on a database of 500 miscellaneous images using this method is illustrated in Figure 5. Here, a query was formed by visually selecting the texture elements corresponding regularly spaced horizontal lines. In s-f notation this is equivalent to requesting the presence of a region with predominant vertical s-f energies. As illustrated in Figure 5, the images retrieved from the database contain regions which have regions with dominant vertical s-f information. It can been seen that the texture feature sets capture salient texture information within images which allow images to be retrieved on the



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

Figure 4. Texture extraction from the "Barbara" image, (a) QMF wavelet image ($y_k's$), (b) subband energy images, (c) after thresholding ($b_k's$) (d) labeled textures ($t[m,n]$), (e) texture regions outlined.

basis of texture content.

Figure 5. Partial results of texture query on database of 500 images. Query texture vector = [1 0 0 1 0 0 0 0 0], which defines textures with low and mid-range vertical *s-f*'s.

4. SUMMARY

We presented a new algorithm for extracting, searching for and retrieving visual data from image and video databases using texture. The binary texture feature sets are produced from image spatial-frequency analysis. The texture analysis captures local texture content and extracts arbitrarily shaped regions of texture from the image. Furthermore, the texture feature sets may be extracted directly from *s-f* compressed image data. Finally, the binary texture feature space lends well to efficient indexing which provides fast image searching and retrieval using texture.

REFERENCES

[1] J. Beck, A. Sutter, and R. Ivry. Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing*, 37, 1987.

[2] Alan C. Bovick, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence*, 12(1), January 1990.

[3] S.-F. Chang. Compressed-domain techniques for image/video indexing and manipulation. In *International Conference on Image Processing*, Special Session on Digital Library and Video-On Demand. I.E.E.E., October 1995.

[4] S.-F. Chang and John R. Smith. Extracting multidimensional signal features for content-based visual query. In *Symposium on Visual Communications and Signal Processing*. SPIE, May 1995.

[5] J.-L. Chen and A. Kundu. Rotation and gray scale transform invariant texture identification using wavelet decomposition and hidden markov model. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence*, February, vol. 16, no. 2 1994.

[6] John G. Daugman. Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *I.E.E.E. Transactions on Biomedical Engineering*, January, vol. 36, no. 1 1989.

[7] E. Griffiths and T. Troscianko. Can human texture discrimination be mimicked by a computer model using local fourier analysis. *Spatial Vision*, vol. 6 1992.

[8] Robert M. Haralick. Statistical and structural approaches to texture. *Proceedings of the I.E.E.E.*, May, vol. 67, no. 5 1979.

[9] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, vol. 24, no. 12 1991.

[10] A. Kundu and J.-L. Chen. Texture classification using qmf bank-based subband decomposition. *CVGIP: Graphical Models and Image Processing*, September, vol. 54, no. 5 1992.

[11] F. Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. Technical report, MIT Media Laboratory and Modeling Group, technical Report No. 320, 1994.

[12] Todd Reed and H. Wechsler. Segmentation of textured images and gestalt organization using spatial/spatial-frequency representations. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence*, January, vol. 12, no. 1 1990.

[13] John R. Smith and S.-F. Chang. Texture classification and discrimination in large image databases. In *Proceedings of the I.E.E.E. International Conference on Image Processing (ICIP-94)*. I.E.E.E., November 1994.