M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 320
Appears in: IEEE Trans. on PAMI, vol. 18, no. 7, July 1996, pp. 722-733
This is a revised version of TR No. 320 (March 1995)

Periodicity, directionality, and randomness: Wold features for image modeling and retrieval*

F. Liu and R. W. Picard
Media Laboratory, E15-390
Massachusetts Institute of Technology, Cambridge, MA 02139
fliu@media.mit.edu, picard@media.mit.edu

Abstract

One of the fundamental challenges in pattern recognition is choosing a set of features appropriate to a class of problems. In applications such as database retrieval, it is important that image features used in pattern comparison provide good measures of image perceptual similarities.

In this paper, we present an image model with a new set of features that address the challenge of perceptual similarity. The model is based on the 2-D Wold decomposition of homogeneous random fields. The three resulting mutually orthogonal subfields have perceptual properties which can be described as "periodicity", "directionality", and "randomness", approximating what are indicated to be the three most important dimensions of human texture perception. The method presented here improves upon earlier Wold-based models in its tolerance to a variety of local inhomogeneities which arise in natural textures and its invariance under image transformation such as rotation.

An image retrieval algorithm based on the new texture model is presented. Different types of image features are aggregated for similarity comparison by using a Bayesian probabilistic approach. The effectiveness of the Wold model at retrieving perceptually similar natural textures is demonstrated in comparison to that of two other well-known pattern recognition methods. The Wold model appears to offer a perceptually more satisfying measure of pattern similarity while exceeding the performance of these other methods by traditional pattern recognition criteria. Examples of natural scene Wold texture modeling are also presented.

1 Introduction

Current worldwide efforts of digitizing massive archives of image, film, and video have created an immediate demand for automated retrieval systems. Tools assisting search among texture-rich imagery have broad applications in, to name a few, video editing, medical image query, and commodity markets such as carpet, tile, and upholstery.

A retrieval system serves the purpose of saving human users the time and effort of browsing the entire database; hence, it is expected that the retrieved images resemble the visual properties of the prototype pattern provided by the human user. To build such a system, it is important that the features used for pattern comparison are faithful to those used by humans in comparing patterns. Considering image retrieval as a pattern recognition application, we face the difficult problem of choosing a set of features for measuring perceptual similarity.

A human texture perception study conducted by Rao and Lohse [1] has indicated that the three most important perceptual dimensions in natural texture discrimination can be described as "repetitiveness", "directionality", and "granularity and complexity". Hence, it is desirable for a texture-based retrieval system to use modeling features which relate image attributes to these perceptual saliencies. We propose here a set of image features based on the two-dimensional (2-D) Wold decomposition of random fields to capture the properties of human texture perception.

Given a 2-D homogeneous random field, the Wold theory allows it to be decomposed into three mutually orthogonal components. The perceptual properties of these components can be described as "periodicity", "directionality", and "randomness", agreeing closely with the most important dimensions of human texture perception.

The 2-D Wold decomposition has been recently applied to spectral estimation and texture modeling by Francos *et al.* [2][3][4]. In their work, it is assumed that the images are homogeneous random fields and the model designs are based on the actual image decomposition. Although their algorithms performed well on a few texture examples, they are not robust or computationally efficient enough to handle databases where image quantity is large and inhomogeneity abounds.

In this paper, we present a new Wold-based texture model ("Wold model" for short) and its applications to image retrieval in a large texture database and to natural scene representation. Our emphasis is on constructing perceptually important features which can be used for image recognition and similarity comparison. In the model we propose, the Wold features which preserve the perceptual property of the Wold components are extracted without having to decompose each image. The algorithms for image modeling and similarity comparison are also designed to tolerate a variety of local inhomogeneities of textures, as well as transformations such as pattern rotation. The problem of aggregating different types of features for image similarity comparison is resolved by using a Bayesian probabilistic approach.

*This work was supported in part by BT, PLC, Interval Research Corp., and NEC

The effectiveness of the Wold model for natural texture modeling is demonstrated in image retrieval experiments in comparison to the performance of two other well-known pattern recognition methods, namely, the shift-invariant principal component analysis (SPCA) [5] and the multiresolution simultaneous autoregressive (MRSAR) [6] modeling. The Wold model appears to offer a perceptually more satisfying measure of pattern similarity while exceeding the performance of these other methods by traditional pattern recognition criteria.

To illustrate how the Wold features can be used in natural scene representations, an image segmentation algorithm and experimental segmentation and representation results are also presented.

Section 2 contains a brief review of the 2-D Wold decomposition theory and a discussion on its previous applications to texture modeling. Section 3 presents the new Wold-based texture model. Section 4 describes the application of the new model to image retrieval, showing retrieval examples of the Wold, the SPCA and the MRSAR methods in comparison. Section 5 demonstrates the Wold texture modeling of natural scene images. A discussion of the strengths and weaknesses of the new Wold model is in Section 6, followed by the Conclusions.

2 Theory Review and Previous Work

To make the paper self-contained, we provide in this section two major theorems of the 2-D Wold decomposition theory of homogeneous random fields. Extensive presentations and the proofs of the theorems can be found in [2][7].

2.1 Theory Review

Let $\{y(m,n)\}, (m,n) \in \mathcal{Z}^2$ be a real valued, regular, and homogeneous random field. On the 2-D plane, a set of total order and non-symmetric half-plane (NSHP) is defined such that the boundary line of a NSHP is of rational slope. Denote this set by $\mathcal{O}$.

Theorem 1 **(2-D Wold decomposition)** *A homogeneous regular random field $\{y(m,n)\}$ can be represented uniquely by the following decomposition:*

$$y(m,n) = w(m,n) + p(m,n) + g(m,n). \tag{1}$$

Field $\{w(m,n)\}$ is **purely-indeterministic** *and has a moving average (MA) representation*

$$w(m,n) = \sum_{(0,0)\preceq(k,l)} a(k,l)u(m-k,n-l), \tag{2}$$

where $\sum_{(0,0)\preceq(k,l)} a^2(k,l) < \infty$ and $a(0,0)=1$. The innovation field $\{u(m,n)\}$ is white. Field $\{p(m,n)\}$ and $\{g(m,n)\}$ are **deterministic**. *Field $\{p(m,n)\}$ is* **half-plane deterministic**. *Field $\{g(m,n)\}$ is* **generalized evanescent** *and $g(m,n) = \sum_{o\in\mathcal{O}} e_o(m,n)$, where $e_o(m,n)$ is the* **evanescent** *field of $\{y(m,n)\}$ with respect to the total-order and NSHP support $o \in \mathcal{O}$. Fields $\{w(m,n)\}$, $\{p(m,n)\}$, and $\{e_o(m,n)\}, o\in\mathcal{O}$, are mutually orthogonal.*

A homogeneous random field has a spectral distribution function (SDF) in the form of a Fourier-Stieltjes integral. Define all spectral functions on the rectangular region $\left[-\frac{1}{2},\frac{1}{2}\right] \times \left[-\frac{1}{2},\frac{1}{2}\right]$.

Theorem 2 *Let $F_y(\xi,\eta)$ be the SDF of a regular homogeneous random field $\{y(m,n)\}$, and let $F_y^s(\xi,\eta)$ denote the singular part of $F_y(\xi,\eta)$. Let $F_w(\xi,\eta)$, $F_p(\xi,\eta)$, and $F_g(\xi,\eta)$ be the SDF of the purely-indeterministic, the half-plane deterministic, and the generalized evanescent components of $\{y(m,n)\}$. Function $F_y(\xi,\eta)$ can be uniquely represented as*

$$F_y(\xi,\eta) = F_w(\xi,\eta) + F_p(\xi,\eta) + F_g(\xi,\eta) \tag{3}$$

where $F_g(\xi,\eta) = \sum_{o\in\mathcal{O}} F_{e_o}(\xi,\eta)$ and $F_{e_o}(\xi,\eta)$ is the SDF of the evanescent field of $\{y(m,n)\}$ with respect to the total-order and NSHP definition $o \in \mathcal{O}$. Function $F_w(\xi,\eta)$ is absolutely continuous and $F_p(\xi,\eta) + F_g(\xi,\eta) = F_y^s(\xi,\eta)$ is singular with respect to the Lebesgue measure.

By Theorem 2, the decomposition of the purely-indeterministic and the deterministic components of a regular homogeneous random field can be achieved by separating the singular and the absolutely continuous components of the SDF. This is known as **Lebesgue decomposition** [8].

In order to apply the 2-D Wold theory to texture modeling, some approximations on the deterministic random fields were made [3]. A half-plane deterministic field is approximated by a **harmonic** random field, which in the spectral domain appears as the 2-D Dirac δ-functions supported by discrete points. The SDF of an evanescent field is absolutely continuous in one dimension and singular in the orthogonal dimension, appearing as 1-D Dirac δ-functions supported by lines with rational slopes.

As shown in (2), the purely-indeterministic field has a white noise driven MA representation. Under certain conditions usually satisfied in practice, a 2-D autoregressive (AR) representation of this field exists [9].

In the following, we refer to the harmonic, evanescent, and indeterministic components of a random field as the **Wold components**.

2.2 Previous Work on Wold-based Texture Modeling

A Wold-based model can be built by decomposing an image into its Wold components and modeling each of the components separately. The Wold texture models reported to date decompose an image via one of two approaches: 1) the spectral decomposition method — an approximation of Lebesgue decomposition as global thresholding of Fourier spectral magnitudes [2][3]; 2) the maximum likelihood (ML) parameter estimation — estimating Wold parameters in the spatial domain by fitting a high-order AR process, minimizing a cost function, and solving sets of linear equations [10]. The effectiveness of these methods was demonstrated on a few texture examples.

Comparing the two approaches, the ML method provides more accurate estimates, but the spectral decomposition method is computationally more efficient. Furthermore, although the Wold theory assumes the homogeneity of the random field, the principle of Lebesgue decomposition can be applied to textures which are not strictly homogeneous, but whose spectral peaks remain sufficiently structured to be extracted.

Based on the discussions above, it is reasonable to choose the spectral approach as a starting point. Given the spectrum of an image, a decomposition algorithm is expected to identify and extract the spectral frequencies associated

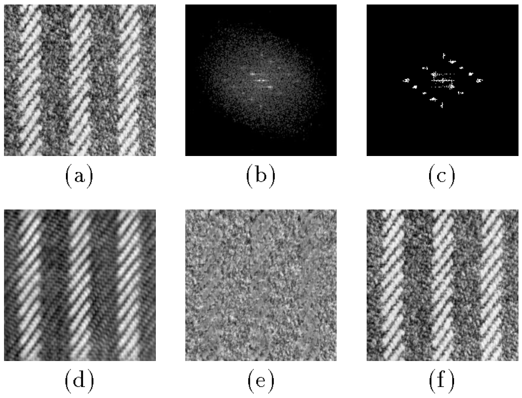


Figure 1: Example of 2-D Wold decomposition. (a) A patch of Brodatz texture D11: Homespun woolen cloth; (b) DFT magnitudes of (a); (c) Identified harmonic frequencies; (d) Extracted harmonic component; (e) Extracted indeterministic component; (f) Synthesized image of (a) from (d) and (e).

with the harmonic and evanescent components. However, we found that a simple global thresholding scheme, as the one used in [3], does not work for many natural textures in the Brodatz database.

A decomposition example is given in Figure 1. A patch of Brodatz texture [11] D11 "Homespun woolen cloth", shown in (a), is decomposed by extracting the harmonic peak frequencies of its Fourier spectrum. Shown in (b), the Fourier magnitude of this pattern has five high frequency peaks[1] which are only locally large. Their values are actually smaller than some of the low indeterministic frequencies. These peaks are important since they give rise to the fine weaving patterns. A global thresholding will either pick up some of the low indeterministic frequencies or omit the weak harmonic peaks. To decompose the texture, two circular Gaussian functions are used to threshold the Fourier magnitudes: one qualifies the peaks and the other determines the region of support. The height and variance of these Gaussians are data dependent. The extracted frequencies, marked in (c), are then removed from the Fourier transform function and inserted into a blank 2-D complex plane. Image (d) and (e) are the inverse Fourier transform of these two complex functions. The synthesized image (f) is the result of adding images (d) and (e).

Although this thresholding scheme is successful in the example, fully automating the decomposition algorithm has been found to be very difficult given the large variety of patterns present in the Brodatz database. Moreover, there are unanswered questions regarding the ambiguity involved in decomposing the values of the harmonic frequencies in discrete spectra. In the past, we have proposed methods of building Wold-based models without actual image decomposition [12][13]. In both methods, textures were classified into harmonic and inharmonic categories and modeled accordingly. However, the binary boundary imposed between the two categories was a major drawback of the approach.

[1] Only half of the 2-D frequency plane is considered due to spectral symmetry.

In the next section, we introduce a new Wold-based model which, without decomposing patterns and setting harsh decision boundaries, extracts and incorporates features that preserve the perceptual property of the Wold components.

3 A New Wold-based Texture Model

3.1 The Construction of the New Model

3.1.1 Brodatz Database

The database used in the retrieval experiments reported in this paper is the "Brodatz texture database". It contains 1008 natural texture patches cropped from all 112 pictures in the Brodatz Album [11]. Each Brodatz texture provides nine 128×128 subimages in 8-bit gray levels. This collection contains a large variety of natural textures, including the many inhomogeneous ones which are not usually included in texture studies. By including the entire Brodatz collection in the database, we allow the potential of confusion and failure that exists when texture algorithms encounter non-texture regions in natural scenes. Examples of the database are shown in Figure 2.

3.1.2 Other Texture Models

Using the benchmarking method reported in [14], the retrieval performance of several image models over the Brodatz database was evaluated by computing their recognition rate operating characteristics. The image classes are defined by the original Brodatz album pages. The average recognition rate (each image in the database is used once as retrieval prototype) is computed for different numbers of the top retrieved images. A 100% recognition rate is reached by a search when 8 matches are found within the top retrieved images considered. For example, if the first 15 retrieved images are considered and 4 matches are found for an image, then the recognition rate for that image at retrieved set size 15 is 50%. The models evaluated include the MRSAR, the SPCA, the tree-structured wavelet transform (TWT) [2] [15], and the three Tamura features of coarseness, contrast, and directionality [16] as used in [17]. Note that this evaluation method uses a traditional pattern recognition criterion, not necessarily agreeing with perceptual criteria.

The benchmarking results show that, when compared to the other three, the MRSAR model offers the best intraclass recognition rate (see Figure 13 in Section 6.2). Recently, a Gabor wavelet decomposition model was also applied to image retrieval and its performance benchmarked against the MRSAR model [18]. By the recognition rate operating characteristics, the retrieval performance of the Gabor and MRSAR methods are similar. Therefore, by this criterion, it would be reasonable to regard MRSAR as a representative of the state-of-the-art texture modeling for image database retrieval. However, in many retrieval cases where structured image patterns are involved, we observe that the MRSAR model is incapable of distinguishing images with very little perceptual resemblance, showing its limitations in measuring perceptual similarity. Examples will be shown in Figure 8 and 9 in Section 4.2. This weakness of the MRSAR model is innate since the model only

[2] The TWT method is sensitive to image sizes. Much smaller than the 512×512 used in [15], the 128×128 database image size had a negative impact on the TWT performance in the benchmarking.

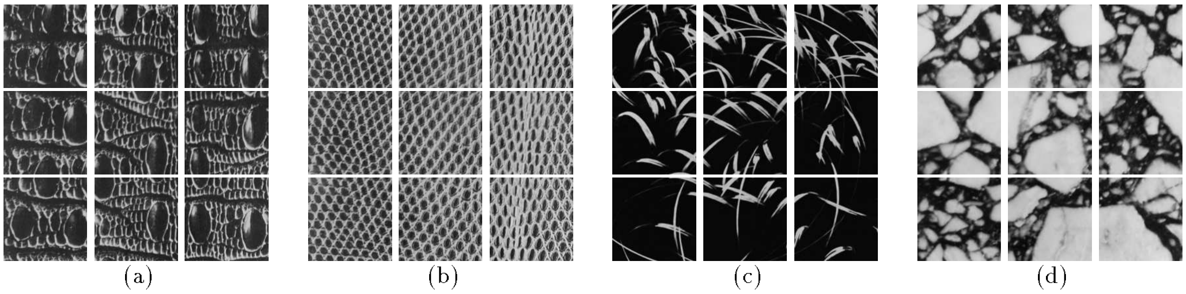


Figure 2: Example images of the Brodatz texture database. (a) D10: Crocodile skin. (b) D36: Lizard skin. (c) D45: Swinging light. (d) D62: European marble.

characterizes the interaction among neighboring image pixels, where neighbors are determined by the model order. As an autoregressive (AR) process, the MRSAR model is most appropriate for modeling random fields with continuous spectra (fine and purely random texture). When using an AR process to model an image with many spectral peaks (spatially periodic structures), it is often difficult to avoid both the information loss inherent in fitting with a low-order model and the extra computation and overfitting with a higher-order model.

3.1.3 Constructing the New Wold Model

Perceptually, by Rao and Lohse's study [1], the existence of periodic structure is the strongest perceptual cue in texture discrimination. We carefully examined the Fourier spectra of all the images in the Brodatz database, concluding the following:

- Natural textures often contain multiple Wold components. Perceptually structured textures usually have dominant harmonic components which appear as structured spectral peaks. Conversely, when the harmonic components are significant, they usually dominate the perceptual pattern discrimination.
- Although certain local inhomogeneities (such as texture on an uneven surface or viewpoint distortion) spread out or change the frequencies of the spectral peaks slightly, the intrinsic structure of these peaks remains.
- Strong evanescent components correspond to eminent directionality in patterns; local inhomogeneities have only a minor effect on these components.

The distinct spectral signatures of some textures from the Brodatz database are shown in Figure 3. The reptile skin in (a) has a prominent harmonic component. The spectral peaks are structured and supported by isolated point-like regions. The cheesecloth in (b) has a strong evanescent component – large spectral peaks supported by a line-like region. The beach sand in (c) is mostly indeterministic, with fairly smooth discrete Fourier transform (DFT) magnitudes.

Considering the observations above and the discussions in Section 2.2, we designed the new Wold-based texture model to first conduct a "harmonicity test" on an image. This test provides a measure of the confidence that the

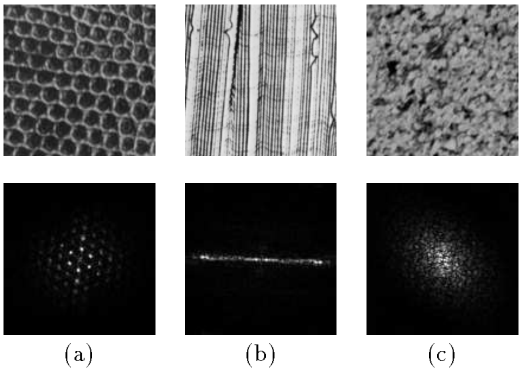


Figure 3: Examples of Brodatz database textures exhibiting distinct spectral signatures in terms of Wold components. Top row: originals; bottom row: DFT magnitudes. (a) D3: Reptile skin, having a prominent harmonic component (spectral peaks supported by discrete points). (b) D105: Cheesecloth, having a strong evanescent component (spectral peaks supported by a line). (c) D29: Beach sand, having mostly a indeterministic component.

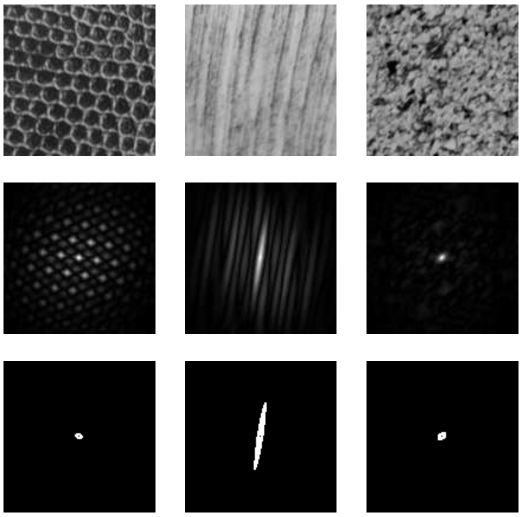
image can be characterized as highly structured (or relatively unstructured). Based on this measure, either harmonic peak feature extraction or MRSAR fitting, or both, are applied. The final Wold representation of the image contains the harmonic confidence measure and the corresponding harmonic peak features and MRSAR features.

The construction of the new model emphasizes the perceptually most salient harmonic information. It also incorporates the demonstrated robustness of the MRSAR model. The new model avoids the decomposition of images. The knowledge of harmonic and indeterministic components is combined probabilistically by using the harmonic confidence measure. Details of the new model are explained in the following subsections.

3.2 "Harmonicity Test" for Feature Extraction

To determine the prominence of harmonic structures in a texture, we examined the energy distribution of image autocovariance functions.

As shown in the top two rows of Figure 4, the autocovari-



(a) (b) (c)

Figure 4: Distinct autocovariance energy distribution of some Brodatz database textures. From top row to bottom: the originals; the absolute value of autocovariance functions; and the small displacement regions. (a) D3: Reptile skin: with periodic energy concentration in the entire displacement plane. (b) D69: Wood grain: with more small displacement energy. (c) D29: Beach sand: with most energy gathered in small displacement region.

ance energy of a highly structured texture is concentrated periodically throughout the 2-D displacement plane. In contrast, the autocovariance energy of a random-looking texture concentrates in a small displacement region. The ratio between the autocovariance "small displacement energy" (defined below) and its total energy (total sum of the absolute value of the function) can be used as an indication of the image harmonicity. (In this work, the autocovariance value at the zero displacement is always ignored.)

An image is first zero-meaned and Gaussian tapered. Standard deviation of 0.375 is used in all image tapering in this work. The image autocovariance is computed as the inverse DFT of the image power spectrum. Starting from the zero displacement, a region is grown outwards continuously until the value of the autocovariance function is lower than a small portion of the function range (10% in our experiments). This region is regarded as the small displacement region. Examples are shown in the bottom row of Figure 4. The energy in this region is used as the "small displacement energy".

The autocovariance energy ratio, r_e, is computed for each image in the Brodatz database. The histogram of these ratios has a bi-modal structure. Gaussian assumptions are made to model the energy ratio data using an expectation and maximization (EM) procedure. Denote the resulting classes as ω_h (harmonic) and ω_r (random). The EM algorithm gives the means and variances of the Gaussian conditional probability density functions of r_e, denoted as $p(r_e|\omega_h)$ and $p(r_e|\omega_r)$, and the prior probabilities, denoted as $p(\omega_h)$ and $p(\omega_r)$. Details of the EM fitting results can be found in Appendix A.

Given the autocovariance energy ratio r_e of an image, the posterior probability of ω_h can be computed as

$$\begin{aligned} p(\omega_h|r_e) &= \frac{p(r_e,\omega_h)}{p(r_e)} = \frac{p(r_e,\omega_h)}{p(r_e,\omega_h)+p(r_e,\omega_r)} \\ &= \frac{p(r_e|\omega_h)p(\omega_h)}{p(r_e|\omega_h)p(\omega_h)+p(r_e|\omega_r)p(\omega_r)}. \end{aligned} \tag{4}$$

This probability is then used as the confidence measure of characterizing the image as highly structured. Consequently, the confidence of describing the image as relatively unstructured is

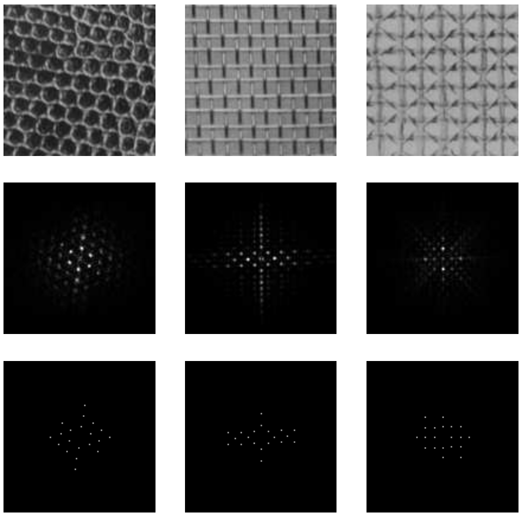
$$p(\omega_r|r_e) = 1 - p(\omega_h|r_e). \tag{5}$$

For a given image, the values of $p(\omega_h|r_e)$ and $p(\omega_r|r_e)$ determine what feature sets are computed. By the property of Gaussian functions, any value of r_e gives non-zero posterior probabilities. To save computation and storage, values of $p(\omega_h|r_e)$ and $p(\omega_r|r_e)$ smaller than 0.001 are considered insignificant and set to zero (for about 5% of Brodatz database images). Corresponding to the non-zero $p(\omega_h|r_e)$ and $p(\omega_r|r_e)$, the harmonic peak features and the MRSAR features are computed respectively.

3.3 Features for Harmonic Structures

The Wold feature set characterizing the harmonic structure of an image consists of the frequencies and the magnitudes of the harmonic spectral peaks. To extract the feature set, the image is first zero-meaned and Gaussian tapered, and then its DFT magnitudes are computed. The local maxima of the magnitudes (excluding values below 5% of the magnitude range) are found by searching a 5×5 neighborhood of each frequency sample. The size of the neighborhood is chosen to match the resolution of the estimated spectra so that the resulting local maxima are separated from each other by at least two frequency samples. Next, the local maxima are examined for the harmonic peaks. A local maximum is a harmonic peak only if its frequency is either a fundamental or a harmonic. A fundamental is defined as a frequency which can be used to linearly express the frequencies of some other local maxima. A harmonic is a frequency which can be expressed as a linear combination of some fundamentals. Starting from the one with the lowest frequency and in ascending order of their frequencies, each local maxima is checked first for its harmonicity — if its frequency can be expressed as a linear combination of the existing fundamentals, and then for its fundamentality — if the multiples of its frequency, combined with the multiples of existing fundamentals, coincide with the frequency of another local maximum. A tolerance of two sample points in both row and column directions is used in the frequency matching. Examples of harmonic Wold features are shown in Figure 5. Note that it is usually not necessary to use all detected harmonic peaks for the feature sets. In this work, only the ten largest ones are kept for each image.

The harmonic Wold features inherit from the Fourier spectral magnitude the property of spatial shift-invariance, a property that is usually important when comparing images. It is often desirable for a retrieval system to also provide users options such as pattern comparison with respect to relative rotation. Local orientation adjustment may also be used to "straighten out" an inhomogeneous



(a) (b) (c)

Figure 5: Harmonic features of three Brodatz database textures. From the top row to the bottom: originals; DFT magnitudes; and harmonic peak feature frequencies. (a) D3: Reptile skin. (b) D14: Woven aluminum wire. (c) D52: Oriental straw cloth.

pattern. Since the spatial relationship of the harmonic peaks in a Wold feature set does not vary under rotation, effects of relative rotation among textures may be reduced by rotating the peaks to align the main orientation of the texture to a chosen direction (horizontal in this work). The main orientation of a texture is defined here as the direction of the lowest fundamental frequency in the feature set. Note that this direction may not correspond to the *perceptually* most salient orientation in the image, but this does not matter for the purposes of comparing images after aligning their orientations. Aligning the peaks using the frequency with the most energy (not necessarily the lowest fundamental frequency) is not as useful since the energy distribution can be influenced by many non-pattern attributes, such as local lighting and contrast. Since each feature set typically consists of a small number of peaks, its rotation involves minimal computation compared to a rotation in the spatial domain. Note that similar savings can be gained on other coordinate transformations.

In image retrieval, the user selects a **prototype image** and the retrieval algorithm searches through the database **test images** for the ones that are similar to the prototype. The comparison of the texture harmonic structures is carried out by matching the Wold feature sets. Denote the peak feature magnitude values of a prototype and a test image by $m_p(s)$ and $m_t(r)$ respectively, where $s = (s_1, s_2), r = (r_1, r_2) \in \mathcal{T}$. Region $\mathcal{T}$ is half of the discrete frequency plane. The harmonic pattern similarity between the two images is measured as:

$$M_{pt} = \sum_{s \in \mathcal{T}} m_p(s) \sum_{r \in \mathcal{T}} w_p(r - s) \frac{m_p(s) m_t(r)}{[m_p(s) + m_t(r)]^2}, \qquad (6)$$

where $w_p(\cdot)$ is a point spread weighting function, implemented here as a 5×5 (size found empirically) Gaussian mask with unity at the center and standard deviation $\sigma = \sqrt{5}$. This function enables peak matching within a small neighborhood of the prototype peaks. This not only compensates for the frequency sampling effects of the DFT operation, but also tolerates small frequency shifts of the harmonic peaks caused by inhomogeneities in the data. The function of the ratio term is to weigh the difference of the peak magnitudes since $\left[\frac{m_p(s)}{m_p(s)+m_t(r)} \cdot \frac{m_t(r)}{m_p(s)+m_t(r)}\right]$ reaches its maximum when $m_p(s) = m_t(r)$. Note that the larger the value M_{pt}, the more similar the two harmonic patterns.

3.4 Features for Relatively Unstructured Textures

The indeterministic component of a texture can be modeled by an AR process (Section 2). Various AR implementations have been used in texture modeling. In this work, we use the second order symmetric MRSAR model of Mao and Jain [6].

The least squares error (LSE) method is used to estimate the MRSAR model parameters. Other methods, such as the ML estimation [19] and the 2-D Levinson type algorithm [20], can also be used. It has been shown that under the experimental circumstances similar to this work, the LSE and the ML estimates offer very similar performance [21]. The 2-D Levinson algorithm is especially useful when the model order determination is involved in the parameter estimation. Since the MRSAR modeling in this work targets the relatively unstructured patterns in an image, a fixed second-order model is chosen and the LSE estimation is used for its computational simplicity.

At each of the second, third, and fourth resolutions, four SAR coefficients are estimated for an image. These coefficients and the estimation error compose a five-parameter vector. The vectors from three resolutions are then concatenated to form a fifteen-parameter MRSAR feature vector. The covariance matrix of the feature vectors within each image is also computed, and two images are compared by examining the Mahalanobis distance of their MRSAR feature vectors. The results of image retrieval based solely on the MRSAR features is compared in Section 4.2 to the performance of the new Wold model.

3.5 Detecting Evanescent Components

Since the spectral signatures of evanescent components are straight lines, an algorithm using the gray-scale Hough transform was developed to detect evanescent components in the frequency domain. After computing the Hough transform of the image DFT magnitudes, the histogram of line slope angles is built. The variance of this histogram and the variance of the Fourier energy along lines corresponding to the sharp peaks in the histogram are found to be discriminative features for evanescent detection. This algorithm accurately identifies the images from the Brodatz pictures D49, D105, and D106 as highly evanescent. Perceptually, these images indeed have distinctively strong directional properties.

The fact that the Brodatz database contains few strongly evanescent samples makes it impossible to statistically determine how the evanescent information should be incorporated into the modeling procedure. In this work, the evanescent database images are modeled by MRSAR processes.

3.6 Measuring Similarity of Textures

Using the Wold features of textures, the image similarities can be measured by either the harmonic peak matching or the MRSAR feature Mahalanobis distances. However, since the harmonic and the MRSAR features are of different types, it is an open question how the two measures should be best combined so that the resulting measure reflects the overall similarity of textures. In the context of image retrieval, we propose the following probabilistic joint measure for image similarity.

Given a prototype image, the system generates two image orderings by using the harmonic peak and the MRSAR features respectively. In each ordering, the entire database is sorted by the descending order of the image similarity to the prototype. For an arbitrary test image, its order numbers in the two orderings are typically different. Denote its order number in the harmonic ordering by O_h and the one in the MRSAR ordering by O_r[3]. As discussed in Section 3.2, we consider the posterior probabilities $p(\omega_h|r_e)$ and $p(\omega_r|r_e)$ as a measure of our confidence to characterize the *prototype* texture as highly structured or relatively unstructured. More specifically, these probabilities indicate the degree of our belief in the two orderings. Hence, the joint order number of the test image is computed as

$$O_{joint} = O_h p(\omega_h|r_e) + O_r p(\omega_r|r_e).$$

The final similarity ordering of the database is formed by sorting images in the ascending order of their joint order numbers.

As an additional benefit, with this similarity measure, we have found that the system is less sensitive to the choices of threshold parameters (such as the 10% for the small displacement energy calculation in Section 3.2), while giving improved overall retrieval performance.

4 Image Retrieval Using the New Wold Texture Model [4]

4.1 Image Retrieval System

The image retrieval algorithm proposed here consists of four stages. The first stage is the harmonicity test and evanescent detection. Given a prototype image, its autocovariance energy ratio is computed to obtain the posterior probabilities $p(\omega_h|r_e)$ and $p(\omega_r|r_e)$. Probability values smaller than 0.001 are set to zero. In the second stage, corresponding to the non-zero posterior probabilities, the harmonic peak feature set and the MRSAR features are estimated respectively. The harmonic peaks in the feature set are rotated to align their main orientation to horizontal. The third stage provides database image orderings where the entire database is sorted by the descending order of the image similarity to the prototype. In each ordering, the similarities are measured by either the harmonic peak matching or the MRSAR feature Mahalanobis distances. In the last stage, different orderings are combined using the method described in Section 3.6 to form the final joint ordering. The flow-chart of this image retrieval system is shown in Figure 6.

0.4	0.6	0.7	0.6	0.4
0.6	0.8	0.9	0.8	0.6
0.7	0.9	1.0	0.9	0.7
0.6	0.8	0.9	0.8	0.6
0.4	0.6	0.7	0.6	0.4

Figure 7: The non-zero region of the Gaussian weighting function $w_p(\cdot)$ for harmonic peak matching ($\sigma = \sqrt{5}$).

The retrieval experiments are carried out on the Brodatz texture database using the Photobook test environment described in [22]. Parameters used to compute the posterior probabilities for a prototype image in harmonicity testing can be found in Table 1 of Appendix A. The Gaussian weighting function for harmonic peak matching is shown in Figure 7.

Each harmonic peak feature set contains the 2-D frequencies and magnitudes of ten harmonic peaks, yielding twenty integers and ten floating-point numbers per image. A MRSAR feature set includes the 15-parameter feature vector and the 15×15 feature covariance matrix (120 distinct numbers due to symmetry), for a total of 135 floating-point numbers per image. For a 128×128 image, feature computation takes typically 0.18 second for the harmonic peaks and 38 seconds for the MRSAR features on an HP9000/735 workstation. The memory needed to store the features of the entire database is 81 kilobytes for the SPCA, 545 kilobytes for the MRSAR, and 580 kilobytes for the Wold.

4.2 Image Retrieval Examples

In Figures 8 and 9, two examples of Wold-based image retrieval are shown together with the results given by the SPCA model and the MRSAR model. The two latter models are described and benchmarked in [14]. The pictures are in the format of the "Photobook" display window. The upper left image is the user selected prototype image. In raster-scan order after the prototype, the retrieved images are shown by descending similarity to the prototype[5]. With pre-computation of the features, all three methods search the database in interactive-time (the search is faster than loading the images for display).

In our experiments, two performance criteria are considered. One is quantitative: the nine samples from each original Brodatz texture form a class and a perfect "traditional pattern recognition performance" implies that all nine images appear at the first row of the output display. The other criterion is qualitative and more difficult to evaluate: the retrieved images should be in the order of their

[3]Images having the same similarity measure value to the prototype share the same order number.

[4]Both the Photobook system and the Wold model software are available by request.

[5]The drawback of sequential display is that images having the same order number appear as different in their ordering.

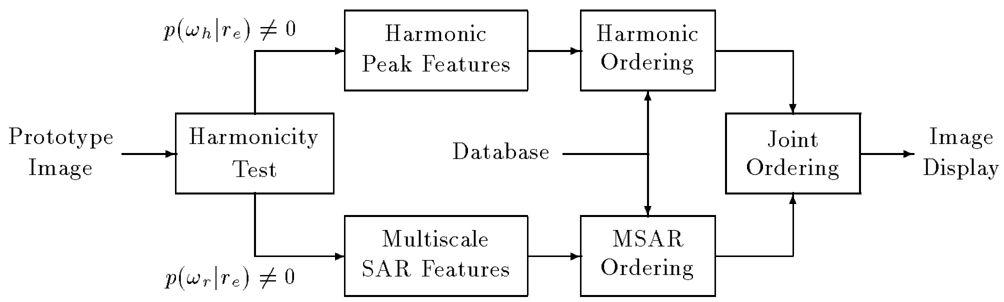


Figure 6: Flow-chart of the image retrieval algorithm based on the new Wold texture model.

perceptual similarity to the prototype image. In fact, the latter criterion is subject to cognitive and other influences. It is not clear if there exists a unique "correct" ordering agreed upon by all people. Our claims about the Wold features being perceptual rely on the studies of Rao and Lohse, and on personal interaction with the features while searching through the Brodatz database.

An example demonstrating the superior qualitative and quantitative performance of the new Wold features is shown in Figure 8. Here, the prototype image is straw cloth. In (a) and (b), both the SPCA and the MRSAR methods fail to find other straw cloth pictures as the most similar; they each retrieve images perceptually very different from the prototype. In (c), the new Wold model provides both "intra-class" accuracy and "inter-class" similarity. It perfectly finds all eight other straw cloth patterns in the database and fills the display with other highly structured textures.

In Figure 9, the experiment is repeated on a prototype image of reptile skin. The results in (a) and (b) show that the SPCA and the MRSAR methods confuse the periodic reptile skin patterns and the random-looking cork patterns. In (c), the Wold method not only just retrieves periodic patterns with many reptile skin images up front, but also shows robustness to the rotational and local inhomogeneities of the reptile skin.

In both examples, the Wold-based method uses largely the harmonic information in the textures ($r_e = 6.36\%$ and 6.41%, $p(\omega_h|r_e) = 0.893$ and 0.892). This is consistent with the fact that both prototype images contain prominent periodic structures.

5 Wold Features and Natural Scene Representation

In this section, we demonstrate how to generate descriptions for textured regions of natural scenes in terms of Wold features. The scene image is first segmented by using its MRSAR features and a K-means-based clustering algorithm. The Wold features are then extracted for the segmented image patches.

5.1 Textured Region Segmentation

Numerous image segmentation methods have been proposed for various tasks [23][24][25]. While the common practice is to partition the entire image, our focus here is to detect and segment sizable and relatively homogeneous regions in a scene. Note that precision of region boundaries is not an important concern in representing natural scene contents for retrieval; it is more important to extract features that provide a basis for subsequent content identification.

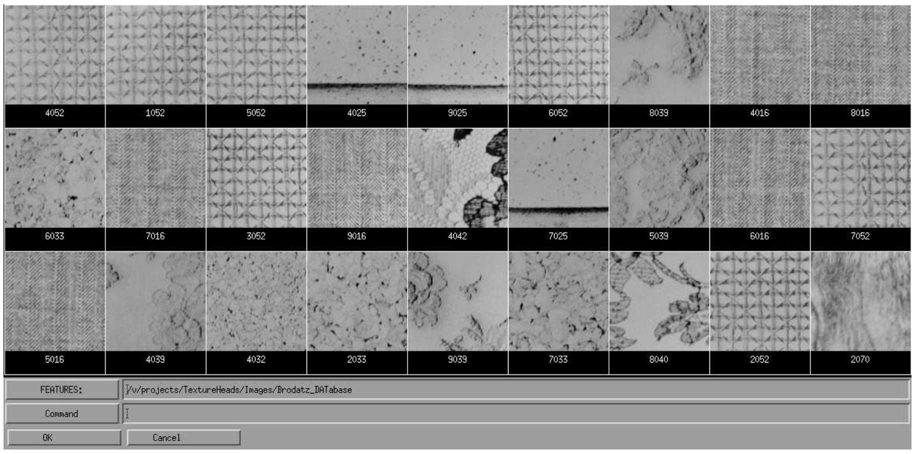
An unsupervised segmentation algorithm has been developed to find reasonably homogeneous image regions. The algorithm is robust to slight inhomogeneities due to perspective viewpoint and uneven textured surfaces. Smooth regions (small variations in pixel values) are first detected by thresholding the local variances at each pixel in a 9×9 neighborhood. These regions are useful for retrieval requests such as "find pictures with a patch of sky at upper left". The main segmentation algorithm is a K-means-based clustering of image pixels in MRSAR feature space. Pixels in smooth regions are excluded from this procedure since the LSE estimates of their MRSAR coefficients are unreliable due to the underdetermined linear equations.

The pixel MRSAR features are the same as described in Section 3.4. To initialize the clustering algorithm, the image is tessellated into rectangular regions (64×64 squares on 256×384 8-bit gray scale images in the experiments below). In a typical iteration, the Mahalanobis distances of each pixel to every cluster are computed and the pixel is re-assigned to the nearest cluster. Small clusters (less than 4000 pixels) are eliminated and their members re-assigned. Clusters are merged when their mutual Mahalanobis distance is small. The program terminates after a given number of iterations or when no pixel changes its cluster membership in an iteration. One morphological closing [26] operation is applied to the segmentation output to smooth the boundaries. The structuring elements used in the two examples shown below have diameters of 15 and 30 pixels respectively.

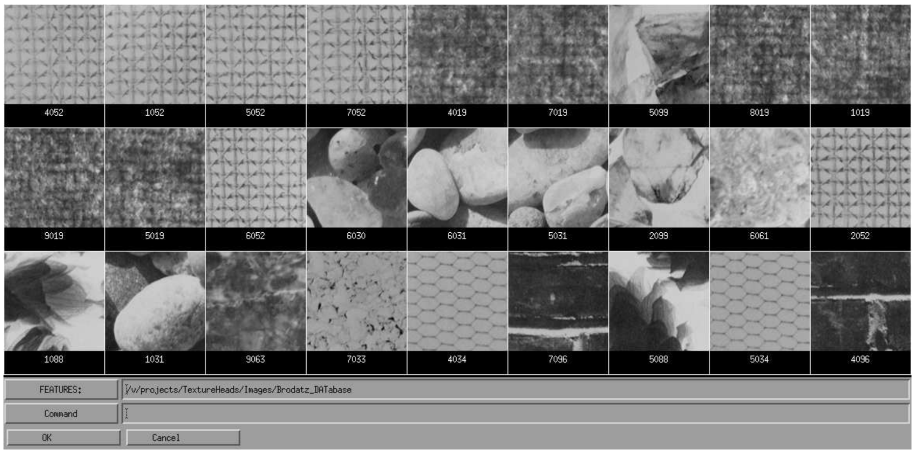
5.2 Natural Scene Representation Examples

Figures 10 and 11 show two examples of textured region segmentation and representation in natural scenes. In both figures, the K-means-based segmentation results are shown with smooth regions marked in black. The number of iterations used in clustering are 15 and 30 respectively for the two images.

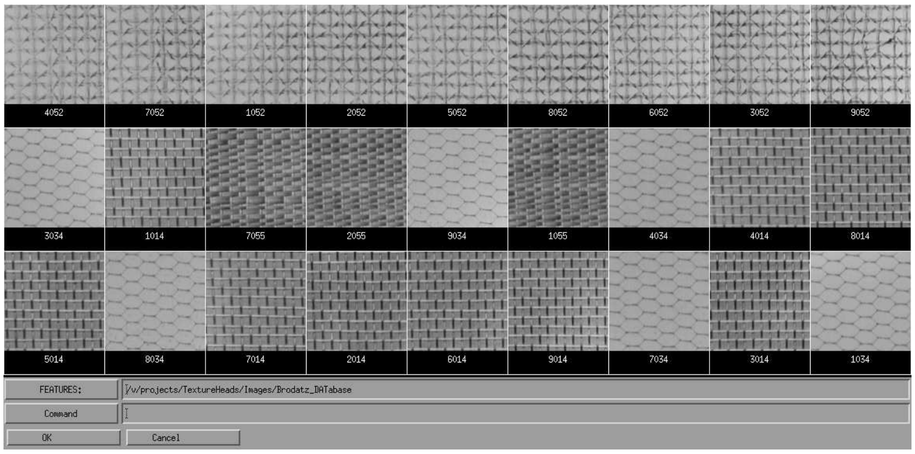
The example shown in Figure 10 illustrates the segmentation and representation of a city scene. The segmented building is shown in (c). The autocovariance energy ratio of



(a)

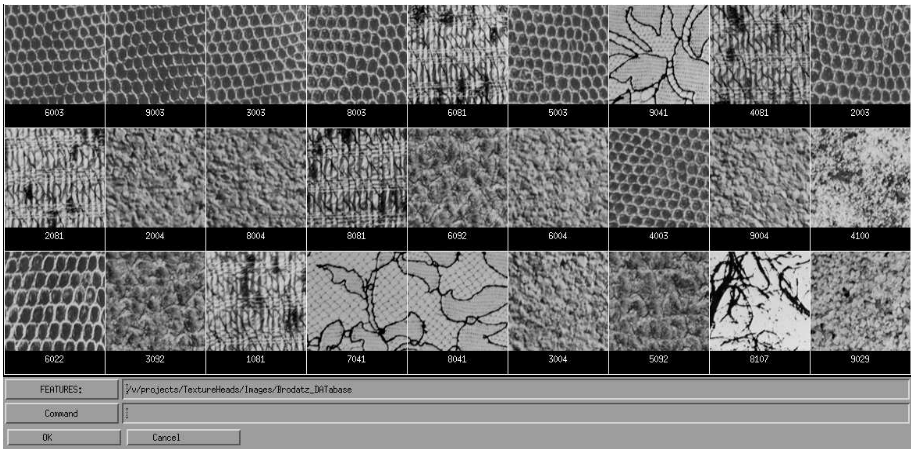


(b)

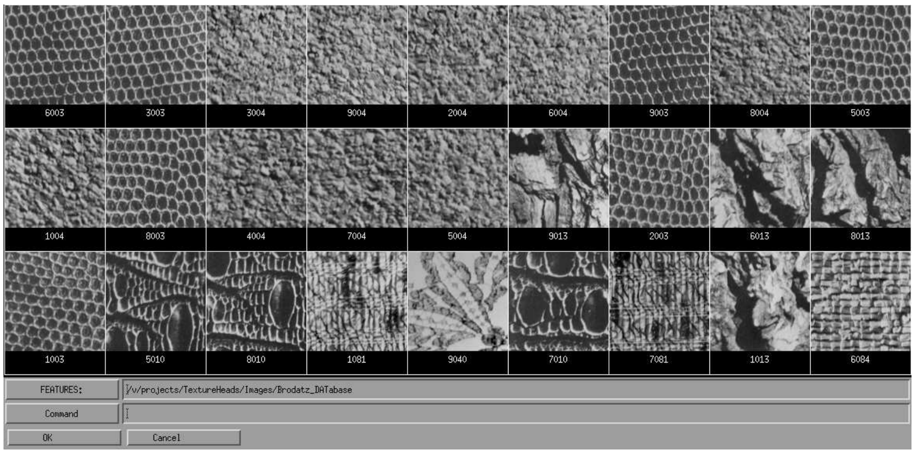


(c)

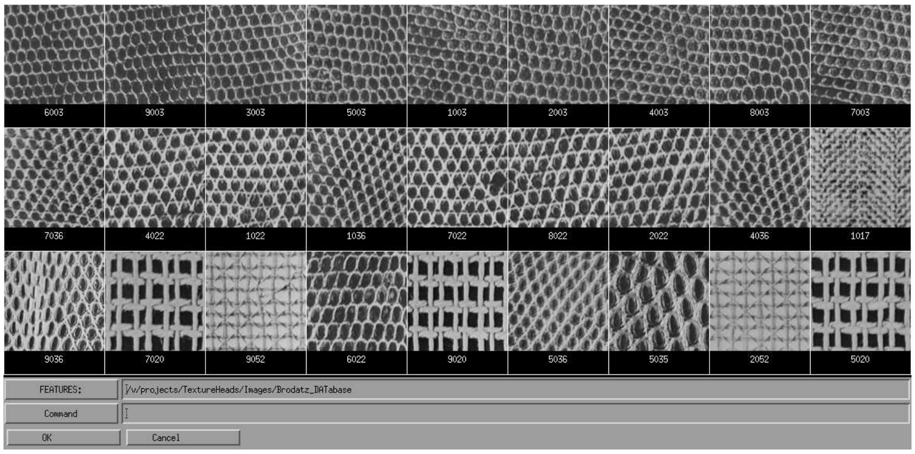
Figure 8: Image retrieval of the straw cloth pattern comparing three methods: (a) SPCA, (b) MRSAR, and (c) Wold. In each picture, the images are raster-scan ordered by their similarities to the image in upper left.



(a)



(b)



(c)

Figure 9: Image retrieval of the reptile skin pattern comparing three methods: (a) SPCA, (b) MRSAR, and (c) Wold. In each picture, the images are raster-scan ordered by their similarities to the image in upper left.

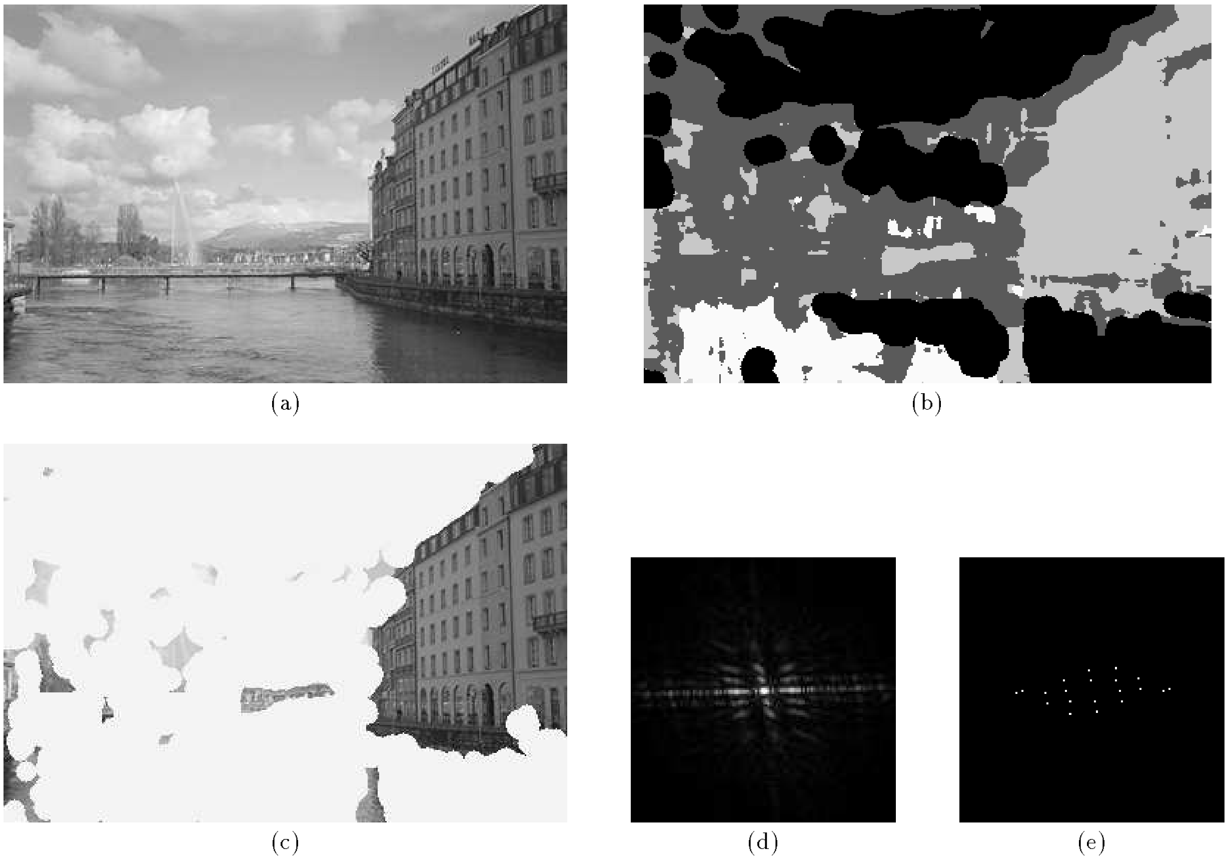


Figure 10: Segmentation of a city scene. (a) Original; (b) Segmentation result with smooth regions in black; (c) Segmented building; (d) DFT magnitudes of building; (e) Extracted harmonic peaks.

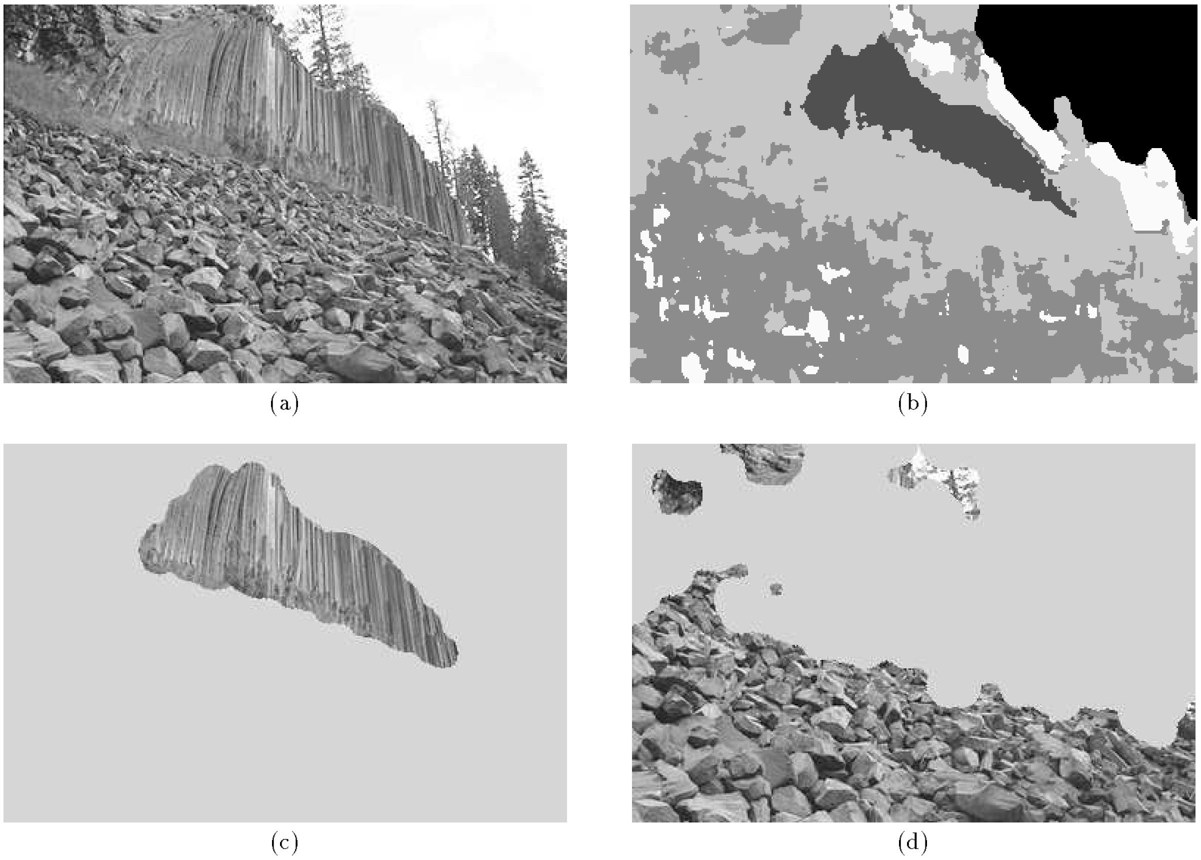
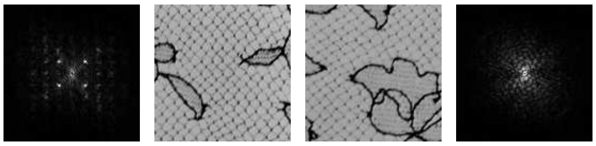


Figure 11: Segmentation of a national park scene. (a) Original; (b) Segmentation result with smooth regions in black; (c) Segmented cliff; (d) segmented rocks.



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

Figure 12: Examples of perceptually similar textures which exhibit distinct spectral signatures. (b) and (c): two patches of Brodatz database texture D41: Lace. (a) and (d): DFT magnitudes of (b) and (c) respectively.

the building is $r_e = 11.88\%$ ($p(\omega_h|r_e) = 0.489$), hence the region should be represented by both the harmonic peak and the MRSAR features. The DFT magnitudes and the harmonic peak features of the building patch are shown in (d) and (e). In computing the DFT, a 128×128 Gaussian window ($\sigma = 24$) is applied to the center of the building. The harmonic peak extraction in this example shows the robustness of the algorithm to inhomogeneity due to perspective, even though no explicit perspective coordinate transform was included.

Note that not only does the segmentation find the building patch in the image, but also the Wold representation of the patch indicates the presence of a "highly structured region." For recognition and retrieval, this description rules out large categories of content such as "grass." If a user were browsing for city scenes, the algorithm could skip over images without any highly-structured regions.

Figure 11 shows a national park scene and its segmentation. Both the cliff and rock patches have no harmonic structures ($r_e > 45\%$) and hence are modeled by their MRSAR features. In addition, the cliff has a strong evanescent component which can be detected by the method described in Section 3.5.

6 Discussion

6.1 Image Inhomogeneity

The effectiveness of the new Wold-based model depends on the properties of the estimated image spectra. On the spectra estimated by the simple method (windowed periodogram) used here, the proposed image modeling and comparison system is surprisingly insensitive to small surface inhomogeneities and viewpoint changes. However, the performance of the new model will be compromised when the inhomogeneities alter image spectra substantially.

One example is given in Figure 12. If shown to a human, the two lace pictures could be judged similar. Nevertheless, one of the lace patterns has prominent spectral harmonic peaks and the other does not. The reason is that the netting pattern in (c) is not homogeneous enough to form strong peaks in its spectra, nor does the netting cover enough area of the image to reinforce the weak periodicity that is present. Instead, the high contrast flowers in (c) overwhelm the harmonic component in spectra. However, a human viewer seems to "homogenize" the netting, and considers the two lace pictures to be similar.

A related issue is the effect of spatial perspective transformations on image spectral peaks. An example is the building patch in Figure 10. Here the algorithm is able to cope with the local inhomogeneity to extract the deformed harmonic peaks. Both theoretical and algorithmic solutions for Wold-based texture modeling under perspective transformation are planned as future research.

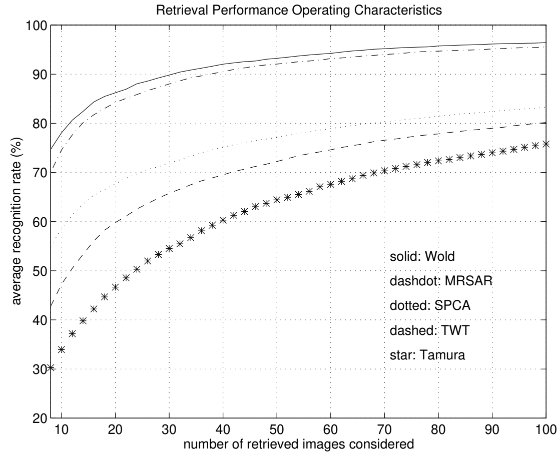


Figure 13: Retrieval performance operating characteristics — average recognition rates over the entire Brodatz database, considering from 8 up to 100 top retrieved images. Methods shown (from top curve to bottom): new Wold-based model, multiresolution simultaneous autoregressive (MRSAR) model, shift-invariant principal component analysis (SPCA), tree-structured wavelet transform (TWT), and Tamura features.

6.2 Performance Over the Brodatz Database

The quantitative measure of the retrieval performance is obtained using the benchmarking method described in 3.1.2. In Figure 13, the average recognition rate characteristics of the new Wold-based method over the Brodatz database is plotted against those of four other methods: MRSAR, SPCA, TWT, and the Tamura features. Figure 13 shows that, by the traditional pattern recognition criterion, the new Wold model is a little more effective than the MRSAR and much better than the SPCA, the TWT, and the Tamura features. However, the key advantage of Wold over the others is not captured by this traditional quantitative benchmark, but involves qualitative searching for perceptually similar patterns, as demonstrated by some examples in the last section.

Comparing the recognition rate averaged within each Brodatz class, the MRSAR method performs better on 10 of the 112 classes at neighbor size 8. Examples are D38 (Water), D41 (Lace), D80 (Straw cloth), and D84 (Raffia). In all ten classes, some patches are quite homogeneous and have prominent spectral peaks while others do not. For the homogeneous patches, the Wold model uses the harmonic information and considers other database images with strong harmonic components as more similar to the prototype than some of the patches in the same Brodatz class. In these classes, the MRSAR model captures the average local spatial interaction and outperforms the Wold model by up to 18%. However, it is arguable if human would agree with the original Brodatz grouping for some of these classes.

Although the Brodatz collection presents the largest variety of natural textures in the research literature to date, it is a very limited set. For instance, the Wold model which represents both the harmonic structure and the overall randomness in a pattern should outperform the MRSAR model on textures with mixed-spectra. However, most of the highly structured Brodatz textures have uniform backgrounds and simple local features. On these images, the MRSAR model, which is incapable of representing large scale spatial arrangements, performs no worse than the Wold model and can even achieve 100% recognition (eg.: D14, D20, D34, and D47). Although the Wold model does better in cases such as D3 (93.1% vs. 54.2%) and D52 (98.6% vs. 58.3%), its strength is not shown strongly against the MRSAR model given the limited variety of the database images.

The fundamental weakness of this performance evaluation is the lack of a meaningful benchmarking method for perceptual similarity in image retrieval. The current classes are defined by the image origin in the Brodatz album, but not the visual similarities. This is especially problematic for inhomogeneous images, where members of different classes can be perceptually more similar than samples from the same original Brodatz picture. Examples are the subimages of D36 (Figure 2 (b) Lizard skin) and D3 (top image of Figure 3 (a) Reptile skin). However, to attempt to regroup the images by their perceptual categories is not as easy as it appears. For example, although semantic grouping is not valid, it is often unnatural for a human to perceive the visual and the semantic similarities among images separately. One example is the five Brodatz brick wall patterns which differ in scale and surface properties. The Wold model will not retrieve all the brick images together and therefore cannot be claimed to be "semantic," although its features do yield semantic descriptions (periodicity, directionality, randomness). The current claims for performance regarding perceptual similarity rely on the relation of the model features to the perceptual dimensions identified by Rao and Lohse.

7 Conclusions

A new texture model based on the 2-D Wold random field decomposition theory is presented and applied to image retrieval in the Brodatz texture database.

The structure of the new model reflects the correspondence between the perceptual properties of the Wold components and the properties of human texture perception. It emphasizes the perceptually most salient harmonic structures in a texture while using the robust statistical models to represent the relatively unstructured patterns. The new model avoids the actual decomposition of images and is designed to tolerate a variety of inhomogeneities in natural data, making it suitable for use in large collections of natural patterns.

The Wold-based model provides a new approach in modeling textures with mixed-spectra. Since the model uses harmonic peak extraction and MRSAR modeling to target different parts of the spectra, it is able to avoid a common problem found in statistical modeling: the information loss inherent in fitting highly structured textures with a low-order model, or the extra computation and overfitting with a higher-order model.

Based on the Wold texture model, a new image retrieval algorithm is proposed. Different types of image features are aggregated for similarity comparison by using a Bayesian probabilistic approach. Compared to other texture models, the Wold model appears to offer perceptually more satisfying results in the image retrieval experiments while exceeding the performance in recognition by traditional quantitative criteria.

A K-means-based image segmentation method is presented to demonstrate how the Wold-based model can be used to characterize textured regions in natural scenes. The Wold feature sets constructed for these regions can be used subsequently in image content description.

	ω_h	ω_r
mean	6.3579	43.6936
variance	10.0241	295.6701
prior	0.1270	0.8730

Table 1: Parameters of two Gaussian classes fitted to the autocovariance energy ratio data.

8 Acknowledgments

The authors wish to thank J. Francos for many stimulating discussions on the 2-D Wold decomposition theory and its applications, T. Minka for the benchmarking of several texture models, and M. Covell for the digitized natural scenes.

A EM Fitting of Autocovariance Energy Ratio Data

Two Gaussian probability density functions are fitted to the autocovariance energy ratio data generated from the entire database. The function parameters estimated by an EM algorithm are shown in Table 1. The joint probability density functions $p(r_e, \omega_h) = p(r_e|\omega_h)p(\omega_h)$ and $p(r_e, \omega_r) = p(r_e|\omega_r)p(\omega_r)$ are plotted in Figure 14, together with the energy ratio histogram.

References

[1] A. R. Rao and G. L. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. In *IEEE Conf. on Visualization*, pages 220–227, San Jose, CA, Oct. 1993.

[2] J. M. Francos. Orthogonal decompositions of 2-d random fields and their applications in 2-D spectral estimation. In N. K. Bose and C. R. Rao, editors, *Signal Processing and Its Applications*, pages 207–227. North Holland, 1993.

[3] J. M. Francos, A. Zvi Meiri, and B. Porat. A unified texture model based on a 2-D Wold like decomposition. *IEEE T. Sig. Proc.*, pages 2665–2678, August 1993.

[4] R. Sriram, J. M. Francos, and W. A. Pearlman. Texture coding using a Wold decomposition model. In

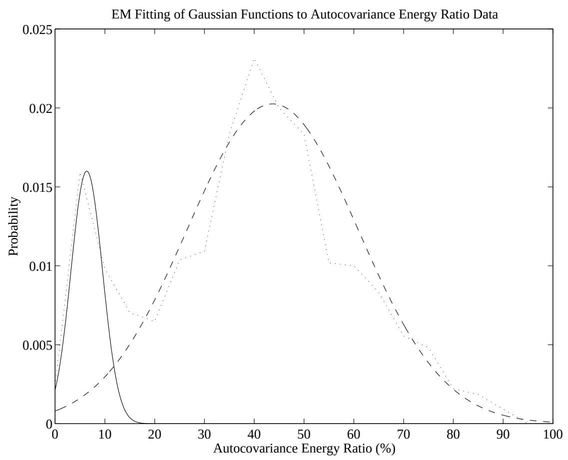


Figure 14: EM fitting of Gaussian functions to image autocovariance energy ratio data. Shown are the joint probability density functions $p(r_e, \omega_h)$ (solid line) and $p(r_e, \omega_r)$ (dash line), together with the energy ratio histogram (dot line). The histogram has bin size 5% and is scaled down by a factor of 5400.

ICPR, volume III, pages 35–39, Jerusalem, Israel, Oct. 1994.

[5] R. W. Picard and T. Kabir. Finding similar patterns in large image databases. In *Proc. IEEE Conf. on Acoustics, Speech, and Signal Proc.*, pages V–161–V–164, Minneapolis, MN, 1993.

[6] J. Mao and A. K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Patt. Rec.*, 25(2):173–188, 1992.

[7] J. M. Francos, A.Z. Meiri, and B. Porat. A Wold-like decomposition of two-dimensional discrete homogeneous random fields. *The Annals of Applied Probability*, 5(1):248–260, Feb. 1995.

[8] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, 1987.

[9] P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434–449, 1954.

[10] J. M. Francos, A. Narasimhan, and J. W. Woods. Maximum likelihood parameter estimation of textures using a Wold-decomposition based model. *IEEE T. Image Proc.*, pages 1655–1666, December 1995.

[11] P. Brodatz. *Textures: A Photographic Album for Artists and Designers.* Dover, New York, 1966.

[12] R. W. Picard and F. Liu. A new Wold ordering for image similarity. In *Proc. IEEE Conf. on Acoustics, Speech, and Signal Proc.*, pages V–129–V–132, Adelaide, Australia, April 1994.

[13] F. Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for perceptual pattern recognition. In *Proc. Int. Conf. Pat. Rec.*, volume II, pages 184–185, Jerusalem, Israel, Oct. 1994.

[14] R. W. Picard, T. Kabir, and F. Liu. Real-time recognition with the entire Brodatz texture database. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 638–639, New York, June 1993.

[15] T. Chang and C.-C. J. Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE T. Image Proc.*, 2(4):429–441, Oct. 1993.

[16] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *IEEE T. Sys., Man and Cyber.*, SMC-8(6):460–473, 1978.

[17] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC project: Querying images by content using color, texture, and shape. In W. Niblack, editor, *Storage and Retrieval for Image and Video Databases*, pages 173–181, San Jose, CA, Feb. 1993. SPIE.

[18] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *IEEE T. Patt. Analy. and Mach. Intell.*, 1996. To appear in special issue on Digital Libraries.

[19] R. L. Kashyap and R. Chellappa. Estimation and choice of neighbors in spatial-interaction models of images. *IEEE T. Info. Theory*, IT-29(1):60–72, 1983.

[20] T. L. Marzetta. Two-dimensional linear prediction: Autocorrelation arrays, minimum-phase prediction error filters, and reflection coefficient arrays. *IEEE T. Acoust., Sp., and Sig, Proc.*, ASSP-28(6):725–733, December 1980.

[21] A. Khotanzad and J.Y. Chen. Unsupervised segmentation of textured images by edge detection in multidimensional features. *IEEE T. Patt. Analy. and Mach. Intell.*, 11(4):414–421, 1989.

[22] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. *Int. J. of Comp. Vis.*, 1996. To Appear.

[23] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Comp. Vis., Graph., and Img. Proc.*, 29:100–132, 1985.

[24] K. S. Fu and J. K. Mui. A survey on image segmentation. *Pattern Recognition*, 13:3–16, 1981.

[25] A. Rosenfeld and L. S. Davis. Image segmentation and image models. *Proc. IEEE*, 67(5):764–772, 1979.

[26] C. H. Richardson and R. W. Schafer. The symbolic manipulation and analysis of morphological algorithms. In Alan V. Oppenheim and S. Hamid Nawab, editors, *Symbolic and knowledge-based signal processing*, pages 142–172. Prentice Hall, 1992.