M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 293
Proceedings of the 12th International Conference on Pattern Recognition,
Jerusalem, Israel, October 9-13, 1994

Periodicity, directionality, and randomness: Wold features for perceptual pattern recognition *

Fang Liu and Rosalind 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 image retrieval, it is important that features used by the system in pattern comparison provide good measures of "perceptual similarity." We present here a new set of features and an image model based on the three mutually orthogonal components produced by the 2-D Wold decomposition of random fields. These components have visual properties which approximate the three most important perceptual dimensions of human texture perception. The method presented here is different from the existing Wold-based models in that it tolerates certain local inhomogeneities which arise in natural textures and reduces computation for comparison of patterns subjected to transformations such as rotation. An image retrieval algorithm based on the new texture model is presented. The effectiveness of the new Wold features for retrieving perceptually similar natural textures is demonstrated by comparing it to that of other well-known pattern recognition methods. The Wold model appears to offer a perceptually more satisfying measure of pattern similarity.

1 Introduction

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

A human texture perception study conducted by Rao and Lohse [1] indicated that the three most important perceptual dimensions in natural texture discrimination can be described as "repetitiveness", "directionality", and "granularity and complexity". We propose a new set of features based on the two-dimensional (2-D) Wold decomposition as appropriate for capturing these features of perceptual similarity.

The 2-D Wold decomposition is an extension of a 1-D decomposition named after statistician H. Wold. Given a 2-D homogeneous pattern, this theory decomposes it into three mutually orthogonal components. The perceptual properties of the three components can be described as "periodicity", "directionality", and "randomness", agreeing closely to the important dimensions of human texture perception. The 2-D Wold decomposition have been recently applied to spectral estimation and texture modeling by Francos *et al.* [2][3].

In this paper, we present a new Wold-based texture model and its application to image retrieval in large texture databases. The Wold-based texture modeling presented to date in the literature assumes that the images are stationary random fields, and, therefore, the model implementations are not designed to handle the inhomogeneity found in large collections of natural image data. In this work, we address the problem of adapting the Wold model to tolerate local inhomogeneities of textures, as well as transformations such as pattern rotation.

Section 2 contains a brief review of the 2-D Wold decomposition theory and its previous applications to texture modeling. Section 3 presents the new Wold-based texture model and its application to image retrieval. Experimental results comparing the Wold-based model with a shift-invariant principal component analysis (PCA) and a simultaneous autoregressive (SAR) model are shown in Section 4.

2 Background

This brief review of the 2-D Wold decomposition is to provide readers some theoretical background of the work presented in this paper. A more complete presentation of the theory can be found in [4][5][6].

2.1 2-D Wold Decomposition of Homogeneous Random Fields

The structure of a 2-D, discrete, real, and homogeneous random field $\{y(m,n)\}$, $(m,n) \in \mathcal{Z}^2$, can be studied by formulating a linear prediction problem. Let $\hat{y}(m,n)$ be the projection of $y(m,n)$ on the Hilbert space spanned by all the samples in the "past" of (m,n) w.r.t. to the **totally ordered, non-symmetric half-plane (NSHP)**. Then

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

the **innovation field** is $\{u(m,n) = y(m,n) - \hat{y}(m,n)\}$. Field $\{y(m,n)\}$ is **regular** if its innovation field does not vanish. Field $\{y(m,n)\}$ is **purely-indeterministic** if it spans the same Hilbert space as its innovation field. Field $\{y(m,n)\}$ is **deterministic** if its innovation field vanishes.

Notice that the total order and NSHP support is not unique in the 2-D lattice. A family of total order and NSHP supports whose boundary lines are of rational slopes can be defined. Denote this family by $\mathcal{O}$. With respect to each support $o \in \mathcal{O}$, there may exist in a deterministic field an **evanescent** subfield which corresponds to the row-to-row innovations. The linear combination of all these evanescent fields is called a **generalized evanescent** field. When a deterministic field has no innovations w.r.t. any total order and NSHP supports, it is **half-plane deterministic**.

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

$$y(m,n) = w(m,n) + p(m,n) + g(m,n). \qquad (1)$$

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

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

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 $\{p(m,n)\}$ 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 $\{y(m,n)\}$ w.r.t. the total order and NSHP support $o \in \mathcal{O}$. Fields $\{w(m,n)\}$, $\{p(m,n)\}$, $\{g(m,n)\}$, and $\{e_o(m,n)\}, o \in \mathcal{O}$, are mutually orthogonal.

This theorem can be proved by using the Theorem 2 in [4] and the Theorem 6 in [5].

There exists a dual relationship between the 2-D Wold decomposition presented by Theorem 1 and the decomposition of the spectral distribution function of a regular homogeneous random field. 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 spectral distribution function 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 spectral distribution functions 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*

$$F_y(\xi,\eta) = F_w(\xi,\eta) + F_p(\xi,\eta) + F_g(\xi,\eta) \qquad (3)$$

where $F_g(\xi,\eta) = \sum_{o\in\mathcal{O}} F_{e_o}(\xi,\eta)$ and $F_{e_o}(\xi,\eta)$ is the spectral distribution function of the evanescent field of $\{y(m,n)\}$ w.r.t. 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 w.r.t. the Lebesgue measure.

The proof of Theorem 2 can be drawn from the proof of Theorem 2 and Theorem 3 in [6] and Theorem 7 of [5].

By Theorem 2, the decomposition of the deterministic and the purely-indeterministic components of a regular homogeneous random field can be achieved by separating the singular and the absolutely continuous components of the spectral distribution of the random field. This is known as **Lebesgue decomposition** [7]. The orthogonality of the two components allows them to be treated separately.

2.2 Approximations

To apply the 2-D Wold decomposition theory to texture modeling, Francos *et al.* made some approximations on the deterministic random field [5]. 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 spectral distribution function of an evanescent field is absolutely continuous in one dimension and singular in the orthogonal dimension. In the spectral domain, this field appears 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 [8][9].

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

3 New Wold-based Texture Model and Application to Image Retrieval

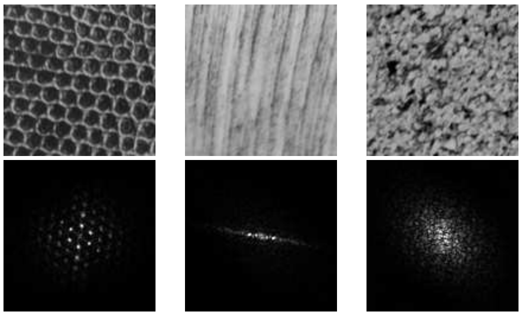
3.1 Construction of the New Model

Image features for retrieval should be able to tolerate certain inhomogeneities in the data while facilitating pattern comparison in real-time. The Wold model implementations in the literature were not designed to meet these requirements; therefore, a new implementation is necessary.

The Wold model implementations reported to date take one of two approaches. One is Lebesgue decomposition [2] [3], and the other is direct maximum likelihood (ML) parameter estimation [10]. Compared to the ML method, the algorithms based on Lebesgue decomposition are computationally more efficient. Also, Fourier spectral analysis has the advantage of being shift-invariant; humans, too, are relatively insensitive to shifts in a texture. Furthermore, although the Wold theory shown previously assumes the homogeneity of the random fields, the principle of Lebesgue decomposition works for textures which are not strictly homogeneous but whose spectral singularities remain structured and can be extracted.

For all the experiments in this paper, we used the "Brodatz texture database" which 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 Brodatz textures which are not usually included in texture studies. We carefully examined the Fourier spectra of this database, drawing the following conclusions:

- Perceptually structured textures usually have dominant harmonic components. Although certain local inhomogeneities may spread out or change the location of the spectral peaks slightly, the intrinsic structure of these peaks remains.



(a) (b) (c)

Figure 1: Examples of Brodatz textures exhibiting different spectral signatures in terms of Wold components. Top row: originals; Bottom row: DFT magnitudes. (a) Reptile skin, having a prominent harmonic component (singularities supported by discrete points). (b) Wood grain, having a strong evanescent component (singularities supported by a line). (c) Beach sand, having mostly indeterministic component.

- Strong evanescent components correspond to eminent directionality in patterns; local inhomogeneities have only a minor effect on these components.
- The spectra of many textures with local inhomogeneity still exhibit structured singularities in the 2-D frequency plane.
- Natural textures may contain multiple Wold components and present a structural to stochastic continuum. However, when the harmonic components are significant, they usually dominate the perceptual pattern discrimination. By Rao and Lohse [1], the existence of periodic structure is the strongest perceptual cue in texture discrimination.

The distinct spectral signatures of some textures from the Brodatz database are shown in Figure 1. The reptile skin in (a) has a prominent harmonic component. The singularities, appear as sharp peaks, locate at isolated point-like regions. The wood grain in (b) has a strong evanescent component – large peaks are over a line-like region. The beach sand in (c) is mostly indeterministic, with fairly smooth discrete Fourier transform (DFT) magnitudes.

Based on the observations above, the new model is implemented with a spectral domain approach. Ideally, the model can be built by applying Lebesgue decomposition directly. However, automating the separation of the singular and continuous parts of the spectrum has been found technically very difficult due to the large variety in the Brodatz texture database. Instead, the new algorithm seeks first the perceptually most salient cue in a texture – the periodicity. When the harmonic information is sufficient, the image is represented by its harmonic peak feature set; otherwise, it is modeled by the number of its main orientations and its multiscale SAR parameters.

3.2 Image Initial Examination

To determine the prominence of harmonic structures in a texture, we examined differences in the energy distribution of the autocorrelation for each image. The autocorrelations of highly structured textures have periodic energy concentration throughout the 2-D displacement plane, whereas those of random looking textures have most of their energy in the small displacement region. Hence, the ratio between the small displacement energy and the total energy of the autocorrelations can indicate if a texture is highly structured.

Each image autocorrelation is computed as the inverse DFT of the image power spectrum. Then, starting from the origin (zero displacement), a region is grown outwards continuously until the value of the autocorrelation function is lower than a small portion of the maximum value (10% in the experiment). This region is regarded as the small displacement region. A histogram of the energy ratio is built over the entire Brodatz database to establish a decision threshold, which is determined to be 18% for minimum classification error (strongly structured vs. not).

3.3 Highly Structured Images and Harmonic Peak features

The Wold feature set of a highly structured texture consists of the frequencies and the magnitudes of the harmonic spectral peaks of the image. To build the feature set, the DFT magnitudes of the image are computed and their large local maxima found. Among the local maxima, only those whose frequencies are either fundamentals or harmonics are kept. Fundamentals are defined as the frequencies which can be used to linearly express the frequencies of other local maxima, and the harmonics are the frequencies which can be expressed as a linear combination of the fundamentals. Note that the feature set usually does not include all the harmonic peaks of a texture.

It is desirable for a recognition algorithm to be able to compare images with respect to relative rotation as humans often consider a texture to be more similar to its rotated version than to a different texture. Local rotations may also be used to "straighten out" an inhomogeneous pattern. Since the spatial relationship of the harmonic peaks in a Wold feature set does not vary under rotation, effects of local inhomogeneities may be reduced by rotating the peaks to align with the main orientation of the texture. This also yields a "generic view" of the texture, analogous to the generic view a human usually draws of an object (such as a building with horizontal and vertical directionality) as opposed to the perspective view formed on the retina.

In the algorithm developed here, the main orientation is defined as the direction of the lowest fundamental frequency in the feature set. The frequency with the most energy is not used since 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, the rotation involves little computation compared to a rotation in the spatial domain. Note that similar savings can be gained on other transformations.

The comparison of highly structured textures is carried out by matching their Wold feature sets. 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. Denote the features of a prototype image 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 similarity measure between the two images is defined in this work to be:

$$M_{pt} = \sum_{s \in \mathcal{T}} m_p(s) \sum_{r \in \mathcal{T}} w_m(r - s) \frac{m_p(s) m_t(r)}{[m_p(s) + m_t(r)]^2}, \quad (4)$$

where $w_m(\cdot)$ is a point spread weighting function, implemented here as a 5×5 (size found heuristically) 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 peaks 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 images.

3.4 Stochastic images and Multiscale SAR Models

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 multiscale second order symmetric SAR model of Mao and Jain [12]. At each of the second, third, and fourth scales, four SAR coefficients are estimated by the least squares error method. These coefficients and the estimation error compose a five-parameter vector. The vectors from three scales are cascaded to form a fifteen-parameter SAR feature vector of each image. The covariance matrix of the feature vector is computed, and two images are compared by examining the Mahalanobis distance of their SAR feature vectors. The results of image retrieval based solely on the SAR features is shown in Section 4 in comparison to the performance of the new system.

The multiscale SAR model is used in this work to model textures without dominant harmonic structures. However, these textures may contain evanescent components which appear in the images as strong directionalities. The evanescent information is described here as the dominant orientations of an image. These orientations are found by using a basis set of oriented bandpass filters and a decision process based on thresholding orientation histograms [13]. The Wold feature set of a texture without prominent harmonic component is composed of the SAR features and the number of dominant orientations in the image. Two textures are compared by computing the Mahalanobis distance of their SAR feature vectors when they have the same number of main orientations.

3.5 Image Retrieval based on the New Wold Texture Model

The image retrieval algorithm proposed here consists of three major parts. Given a prototype image, the system first examines it for strong harmonic structures. When the image is considered highly structured, its harmonic peaks are extracted and rotated to form its feature set. Then, all highly structured database images are sorted by the descending order of their similarity measure (4) to the prototype image. When the prototype image is considered not highly structured, its main orientations and SAR features are estimated to form the feature set. All database images which are not highly structured but possess the same number of main orientations as the prototype image are then sorted by the ascending order of their SAR feature Mahalanobis distances to the prototype image. The flowchart of this Wold-based image retrieval system is shown in Figure 2.

The multiscale SAR parameter estimation is the most computationally costly part of the entire algorithm. However, when the initial stage indicates that the harmonic information is sufficient for the pattern comparison, this operation can be avoided and results in substantial savings. Furthermore, consideration of first the harmonic component and then the other Wold components is consistent with the perceptual saliency of these components [1].

4 Experiments

The retrieval experiments are carried out on the Brodatz texture database using the Photobook test environment described in [14].

In Figure 3, the performance of the new Wold-based texture model is shown by two examples and compared to that of the two other models described and benchmarked in [15]: a shift-invariant PCA model and a multiscale SAR model. The pictures are in the format of the "Photobook" display window. The upper left image is the user selected prototype image and the others are the database images shown in the descending order of their similarities to the prototype image in raster scan.

In our experiments, two performance criteria are considered. One is quantitative: since there are nine samples for each original Brodatz texture, a perfect pattern recognition performance implies that all nine images appear at the first row of the output display. The other is qualitative: the retrieved images should be in the order of their perceptual similarity to the prototype image.

In the left column of Figure 3, the prototype image is of bricks. The results in (a) and (b) show that the shift-invariant PCA method does better in filling the screen with "perceptually similar" images, but the multiscale SAR method is better at finding the other eight brick pictures cropped from the same Brodatz image. In (c), the Wold model combines the best of both – providing both "within-class" accuracy and "inter-class" similarity. Although the brick images tend to appear very structured to humans (who may imagine the periodic wall in which they lie) there is not enough periodicity in the images for the Wold model to find strong harmonics. Hence, the Wold model uses the evanescent and indeterministic components for this case.

In the right column, the experiments are repeated for a prototype image of reptile skin. The experiments in (a) and (b) show that both the shift-invariant PCA and the multiscale SAR methods confuse the periodic reptile skin patterns and the random looking cork patterns. In (c), the Wold method not only retrieves other periodic patterns, but also shows tolerance to the rotational inhomogeneities of the reptile skin. In this example, the Wold method uses only the harmonic information of the textures.

With pre-computation of the features, all three methods above search the database in real-time on a DEC 5000 workstation.

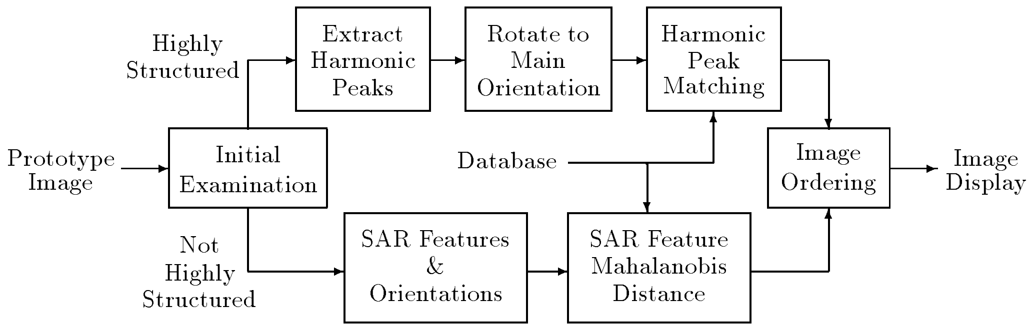


Figure 2: Flow-chart of the image retrieval algorithm based on the new Wold texture model.

5 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 three component fields resulting from the decomposition have perceptual properties associated with the three main perceptual dimensions identified in an independent study of human texture perception.

Adopting the principles of Lebesgue decomposition, the new model represents natural textures by their Wold features – harmonic information for the highly structured textures, and orientation and multiscale SAR features for the relatively unstructured textures. The Wold feature set and the corresponding similarity measuring scheme enable the algorithm to tolerate certain local inhomogeneities in data, making the model more suitable for natural texture modeling.

The new model always seeks first the periodicity information in a texture and represents the image by this information when it is sufficient. This not only avoids the computational burden of fitting to the image a statistical model, but also makes the pattern matching under rotation a simple operation. This modeling procedure is also consistent with the observation that periodicity is the most important perceptual dimension in texture discrimination.

The Wold model also solves a common problem found when trying to fit statistical models to textures with periodic structures (i.e., spectral discontinuities). Since the Wold model treats the continuous and discontinuous spectral components separately, it is able to better fit each component, avoiding the information loss inherent in a low-order model, or the extra computation and overfitting problems in a higher-order model.

Based on the Wold features of images, a new image retrieval algorithm is proposed. The effectiveness of the new model is demonstrated in the image retrieval experiments in comparison to the performance of the shift-invariant PCA model and the multiscale SAR model. The Wold model appears to offer perceptually more satisfying results when applied to the Brodatz textures.

6 Acknowledgments

The authors wish to thank J. Francos for many stimulating discussions on the 2-D Wold decomposition theory and its applications, and M. Gorkani for providing software to estimate orientation in texture.

References

[1] A. R. Rao and G. L. Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. In *IEEE Conf. on Visualization*, San Jose, 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] J.M. Francos, A. Zvi Meiri, and B. Porat. Orthogonal decompositions of 2-D nonhomogeneous discrete random fields. *Submitted For Publication.*

[5] J.M. Francos, A. Zvi Meiri, and B. Porat. A Wold-like decomposition of 2-D discrete homogeneous random fields. *Submitted For Publication.*

[6] H. Helson and D. Lowdenslager. Prediction theory and fourier series in several variables.ii. *Acta Math*, 106:175–213, 1962.

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

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

[9] M. P. Ekstrom and J. W. Woods. Two-dimensional spectral factorization with applications in recursive digital filtering. *IEEE T. Acoust., Sp., and Sig, Proc.*, ASSP-24(2):115–128, April 1976.

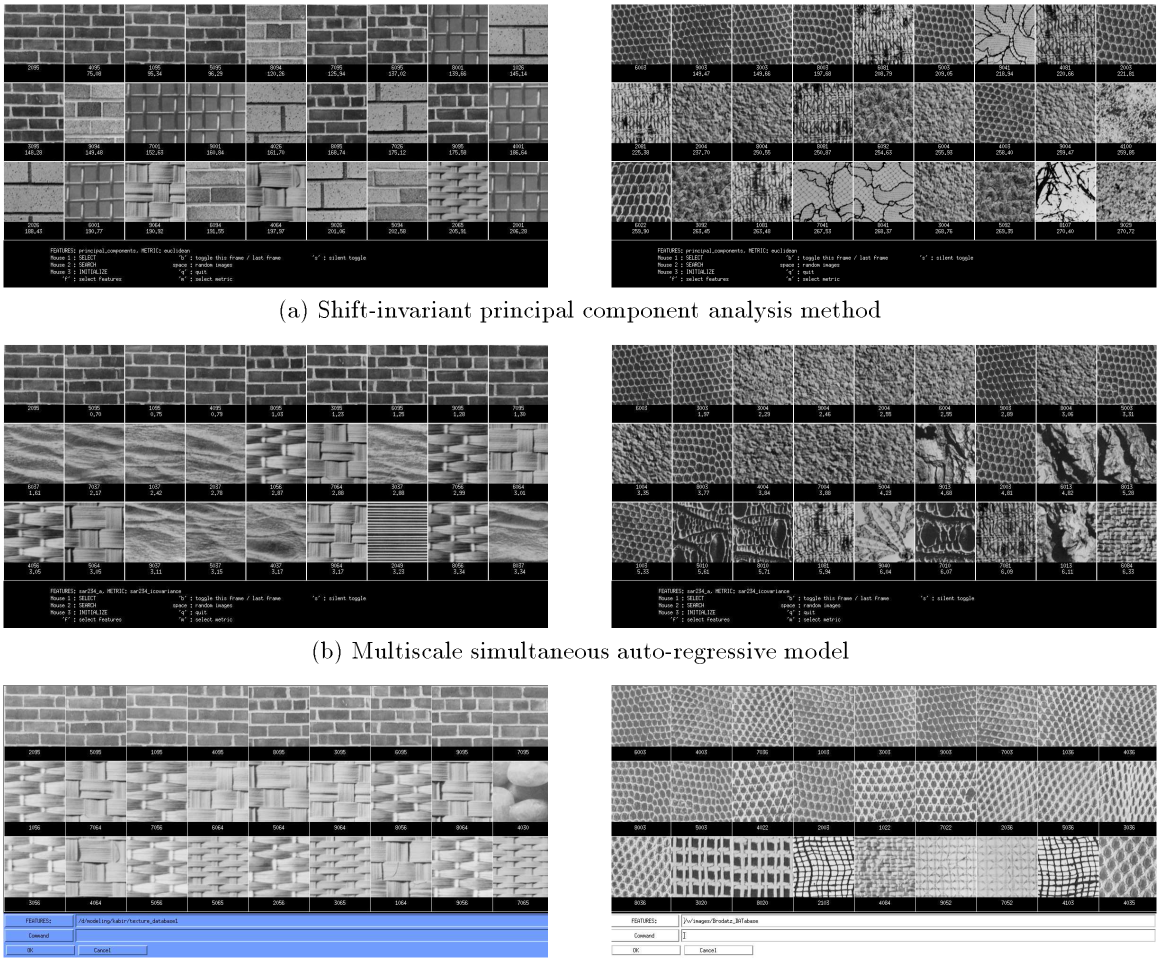
[10] J. M. Francos, A. Narasimhan, and J. W. Woods. Maximum Likelihood Parameter Estimation of the Harmonic, Evanescent and Purely Indeterministic Components of Discrete Homogeneous Random Fields. unpublished.

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

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

[13] R. W. Picard and M. Gorkani. Finding perceptually dominant orientations in natural textures. *Spatial Vision, special Julesz birthday issue.* To Appear; also

D95: Brick wall D3: Reptile skin



(a) Shift-invariant principal component analysis method

(b) Multiscale simultaneous auto-regressive model

(c) New model base on the 2-D Wold decomposition

Figure 3: Image retrieval results comparing three models. (a) Shift-invariant PCA, (b) Multiscale SAR, and (c) new Wold features. In each display, the images are raster-scan ordered by their similarities to the upper left image. Two examples: brick wall (left) and reptile skin (right).

available as Percep. Comp. TR #229, MIT Media Laboratory, 1993.

[14] R. W. Picard and T. Kabir. Finding similar patterns in large image databases. In *Proc. ICASSP*, pages V–161–V–164, Minneapolis, MN, 1993.

[15] 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.