Indexing and Retrieval in Multimedia Libraries Through Parametric Texture Modeling using the 2D Wold Decomposition

Radu Stoica, Josiane Zerubia, Joseph M. Francos

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Indexing and Retrieval in Multimedia Libraries Through Parametric Texture Modeling using the 2D Wold Decomposition

Radu Stoica — Josiane Zerubia — Joseph M. Francos

N° 3594

December 1998

THÈME 3

Rapport de recherche

ISSN 0249-6399

Indexing and Retrieval in Multimedia Libraries Through Parametric Texture Modeling using the 2D Wold Decomposition

Radu Stoica , Josiane Zerubia*, Joseph M. Francos †

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Ariana

Rapport de recherche n° 3594 — December 1998 — 26 pages

Abstract: This paper presents a parametric method for indexing and retrieval of multimedia data in digital libraries. Indexing (labeling) and retrieval of the multimedia data are performed using parametric modeling of the textured segments found in the data imagery components. The estimated parametric models of the textured segments serve as their indices, and hence as indices of the entire image, as well as of the multimedia record which the image is part thereof. To achieve the ability to identify textured image regions and estimate their parameters, a joint segmentation-estimation algorithm that combines the 2-D Wold decomposition based texture model with a Markovian labeling process, is derived. Ordering and indexing of images require a definition of a distance measure between images. Using the framework of the Kullback distance between probability distributions, a new rigorous distance measure between textures is derived. The distance between any two textured image segments is evaluated using their estimated parametric models. The proposed segmentation, distance evaluation, and indexing methods are shown to produce comparable results to those obtained by a human viewer.

Key-words: 2D Wold decomposition, Markov random field, Kullback distance, texture, segmentation, indexing, multimedia.

The work of Radu Stoica and Josiane Zerubia was supported in part by AFIRST (Association Franco-Israelienne pour Recherche Scientifique et Technique) under Grant 960 MRAI 01.

The work of Joseph M. Francos was supported in part by the Israel Ministry of Science and the French Ministry of Research and Technology under Grant 8814196

* Radu Stoica and Josiane Zerubia are with INRIA Sophia Antipolis, 2004 route de Lucioles, BP 93, 06902 Sophia Antipolis Cedex, France.
† Joseph M. Francos is with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel.

Indexation et recherche dans une base de données multimedia grâce à une modélisation paramétrique de texture utilisant la décomposition de Wold 2D

Résumé : Ce rapport présente une méthode paramétrique permettant de faire de l'indexation et de la recherche dans une base de données multimédia. L'indexation (étiquetage) et la recherche de données multimédia sont réalisées grâce à la modélisation paramétrique de textures qui se trouvent dans les images de la base de données. Les textures sont caracterisées par des paramètres qui servent d'indices pour la recherche dans la base de données. Afin de pouvoir identifier les différentes régions texturées d'une image et estimer les paramètres correspondants, un algorithme de segmentation-estimation est proposé dans ce rapport, qui fait appel à une décomposition de Wold 2D pour le modèle de texture et à un modèle markovien pour l'étiquetage. L'indexation nécessite de définir une distance entre les images. Une nouvelle distance, inspirée de la distance de Kullback, est décrite dans ce rapport. Elle utilise les paramètres estimés correspondants au modèle 2D de chaque texture. Les résultats obtenus relativement à la segmentation et à l'indexation sont proches de ceux obtenus par un opérateur humain.

Mots-clés : décomposition de Wold 2D, champ de Markov, distance de Kullback, segmentation, indexation, multimédia.

Contents

1 Introduction

One of the major problems in utilizing the information highways technologies is the need to develop efficient tools for storage, search and navigation in multimedia libraries. More specifically, as the access to multimedia libraries and archives is becoming an integral part of many multimedia applications, efficient tools for storage, search and navigation in multimedia libraries are essential [10, 11]. There is large variety and a broad range of applications like satellite imaging and remote sensing, medical diagnosis and research, film-editing, and advertising, that require users to access very large databases consisting of multimedia data. Such data can be a combination of images, speech and sound, signals, and text. A major problem with multimedia databases is the tremendous volumes of data. A different, yet strongly related problem, is the problem of indexing, labeling, and retrieval of multimedia data. The question of how can one intelligently search in a collection of pictures has no straightforward answer. Yet, due to the tremendously large volumes of data involved, one would like to be able to pinpoint the exact information he, or she, is interested in, to save the time and network resources involved in transferring large volumes of unnecessary data (even in a compressed form).

In this paper we address one aspect of this complex problem, namely, the development of algorithms for indexing (labeling) and retrieval of multimedia data, based on the properties of the imagery components of the stored data record. The proposed solution to the multimedia data retrieval problem is based on content-based retrieval of imagery data. We present tools and techniques to identify images that are similar, or that have a common characteristic with a specified image. More specifically, indexing and retrieval of the data are performed using parametric modeling of the textured components in the imagery data.

In multimedia data retrieval applications, it is important that image features used for pattern comparison provide a good measure of image perceptual similarity. In [16], we have presented a parametric texture model which is based on the 2-D Wold decomposition [14], and corresponding estimation algorithms for the parameters of individual textures. It is well known that texture eludes precise definition. However, in [16] it is shown that the Wold decomposition based texture model is very successful in estimating the parameters of what human viewers perceive as natural textures, and in reproducing the original texture using *only* the estimated parameters. More specifically, using this parametric model it is shown that assuming the texture field is a realization of a Gaussian random field with mixed spectral distribution, essentially indistinguishable replicas of the original texture are synthesized from the estimated parameters. Hence in this paper, texture information is used for characterizing image regions. Using parametric texture modeling we derive an algorithm for labeling and indexing images by jointly segmenting and estimating the parametric models of the textured areas in those images. The parametric model of each textured patch is employed as the *index* of that patch, and therefore as an index to the entire image, and multimedia data record.

Moreover, the parametric representation of a texture, is useful not only as a description of its qualitative features, or as its label, but also for compression purposes. It is shown in [13] that using the Wold-decomposition based texture model, very high compression ratios for textured areas can be achieved. Since the very same parametric model of the texture field is used for coding, storing, and indexing the texture data, the compressed data (*i.e.*, the quantized model parameters) simultaneously provides the (parametric) feature vector that characterizes the texture, and its index. Thus the proposed parametric indexing scheme provides a representation that facilitates content-based retrieval of the compressed data.

The paper is organized as follows: In section 2 a brief description of the Wold decomposition based texture model, is given. The estimation procedure of the texture model parameters is summarized in Section 3. In Section 4, new distance measures for textures are derived. We first present a distance measure that employs the Kullback distance between two multivariate Gaussian probability distributions. This measure has the advantage of naturally incorporating the distances of the different components of the Wold decomposition based random field model, into a single measure. To reduce the complexity of the distance computation, a hierarchical approximation of the distance measure is derived. A texture segmentation procedure that employs the texture parametric model and the new distance measure is presented in Section 5. In Section 6 we present our conclusions.

2 The Texture Model

The presented texture model is based on the results of the Wold-type decomposition of 2-D regular and homogeneous random fields, [14]. Let $\{y(n,m),(n,m) \in \mathcal{Z}^2\}$, be a real valued, regular, and homogeneous random field. Then $y(n,m)$ can be uniquely represented by the orthogonal decomposition

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

The field $\{w(n,m)\}$ is purely indeterministic and has a unique white innovation driven moving average representation, given by

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

The field $\{v(n,m)\}$ is a deterministic random field. It can be shown that there exists a family of NSHP total-order definitions such that the boundary line of the NSHP has a rational slope. Let α, β be two coprime integers, such that $\alpha \neq 0$. The angle θ of the slope is given by tan $\theta = \beta/\alpha$. Each of these supports is called *rational non-symmetrical half-plane* (RNSHP). We denote by O the set of all possible RNSHP definitions on the 2-D lattice, (*i.e.*, the set of all NSHP definitions in which the boundary line of the NSHP is of rational slope). The introduction of the family of RNSHP total-ordering definitions results in the following countably infinite orthogonal decomposition of the field's deterministic component, [14]:

$$v(n,m) = p(n,m) + \sum_{(\alpha,\beta)\in O} e_{(\alpha,\beta)}(n,m) \tag{3}$$

The random field $\{p(n,m)\}$ is a *half-plane deterministic* field, and $\{e_{(\alpha,\beta)}(n,m)\}$ is the evanescent field corresponding to the RNSHP total-ordering definition $(\alpha,\beta) \in O$. Since for practical applications we can exclude singular-continuous spectral distribution functions from the framework of our treatment, a model for the evanescent field which corresponds to the RNSHP defined by $(\alpha, \beta) \in O$ is given by

$$e_{(\alpha,\beta)}(n,m) = \sum_{i=1}^{I^{(\alpha,\beta)}} s_i^{(\alpha,\beta)}(n\alpha - m\beta)\cos\left(2\pi \frac{\nu_i^{(\alpha,\beta)}}{\alpha^2+\beta^2}(n\beta + m\alpha)\right) + t_i^{(\alpha,\beta)}(n\alpha - m\beta)\sin\left(2\pi \frac{\nu_i^{(\alpha,\beta)}}{\alpha^2+\beta^2}(n\beta + m\alpha)\right) \tag{4}$$

where the 1-D purely-indeterministic processes $\{s_i^{(\alpha,\beta)}(n\alpha - m\beta)\}$, $\{s_j^{(\alpha,\beta)}(n\alpha - m\beta)\}$, $\{t_k^{(\alpha,\beta)}(n\alpha - m\beta)\}$, $\{t_\ell^{(\alpha,\beta)}(n\alpha - m\beta)\}$ are mutually orthogonal for all $i, j, k, \ell, i \neq j, k \neq \ell$, and for all i the processes $\{s_i^{(\alpha,\beta)}(n\alpha - m\beta)\}$ and $\{t_i^{(\alpha,\beta)}(n\alpha - m\beta)\}$ have an identical autocorrelation function. Hence, the "spectral density function" of each evanescent field has the form of a countable sum of 1-D delta functions which are supported on lines of rational slope in the 2-D spectral domain. In the following, we assume that the modulating 1-D processes $\{s_i^{(\alpha,\beta)}(n^{(\alpha,\beta)})\}$ and $\{t_i^{(\alpha,\beta)}(n^{(\alpha,\beta)})\}$ of each evanescent field can be modeled by a finite order AR model.

In this paper, we also assume that the half-plane deterministic field consists only of the harmonic random field :

$$h(n,m) = \sum_{p=1}^{P} C_p \cos 2\pi(n\omega_p + m\nu_p) + D_p \sin 2\pi(n\omega_p + m\nu_p) \tag{5}$$

where in general, P is infinite. This component generates the 2-D delta functions of the "spectral density".

When expressed in the general form (5), the coefficients $\{C_p, D_p\}$ of the harmonic component are real valued, mutually orthogonal random variables. However, since in general, only a single realization of the random field is observed we cannot infer anything about the variation of these coefficients over different realizations. The best we can do is to estimate the particular values which the C_p's and D_p's take for the given realization; in other words in the estimation problem we might just as well treat the C_p's and D_p's as unknown constants.

The parametric modeling of deterministic random fields whose spectral measures are concentrated on curves other than lines of rational slope, or discrete points in the frequency plane, is still an open question to the best of our knowledge.

Hence, the observed texture field $\{y(n,m)\}$ is uniquely represented by the orthogonal decomposition $y(n,m) = w(n,m) + h(n,m) + \sum_{(\alpha,\beta)\in O} e_{(\alpha,\beta)}(n,m)$. Thus, the problem of estimating the texture model parameters, becomes one of estimating the parameters of the harmonic and evanescent components of the field in the presence of an unknown colored noise generated by the purely-indeterministic component, jointly with estimating the purely-indeterministic component parameters.

3 Estimation of the Model Parameters

The implementation of a multimedia data retrieval system where data indexing is based on the textural properties of the imagery components of the data records, relies on the existence of a robust estimation algorithm of the texture model parameters. This is since computation of the distance between any two textures requires the substitution of their estimated parametric models into some distance measure.

In [15, 16], we developed a conditional maximum-likelihood algorithm for jointly estimating the parameters of the harmonic, evanescent and purely-indeterministic components of the texture. It is shown that by introducing appropriate parameter transformations the highly nonlinear least-squares (NLLS) problem that results from maximizing the conditional likelihood function is transformed into a separable NLLS problem. Hence, the computational complexity of the required numerical minimization is reduced significantly. In the transformed problem only the spectral support parameters of the deterministic components enter non-linearly into the transformed model equation. Thus, a solution for the unknown spectral supports of the harmonic and evanescent components reduces the problem of solving for the transformed parameters of the field to linear least squares.

Nevertheless, this algorithm still requires the minimization of a non-quadratic cost function with respect to the spectral support parameters of the harmonic and evanescent components. Since this estimator involves a multi-dimensional search in the parameter space it is not practical for a database retrieval application, where short response time is mandatory.

We therefore present in this paper a suboptimal (relative to the maximum likelihood estimator), but computationally efficient algorithm (since no multi-dimensional search in the parameter space is required), for estimating the texture model parameters. The proposed algorithm employs the experimental observation that in natural textures the deterministic component contains harmonic components *or* evanescent components, but not both. This observation enables us to considerably simplify the estimation procedure.

Next, we briefly summarize the suggested algorithm and introduce some necessary notations. The first step of the algorithm is a test for the existence of deterministic components in the observed texture. This test is similar to the one proposed in [17]. Since the deterministic component gives rise to singularities (1-D and 2-D delta-functions) in the spectral domain, its contribution may be evaluated by a thresholding procedure on a spectral density estimate. For the purpose of this test, we use the periodogram, which involves sampling of the spectral domain at discrete points. Therefore, the deterministic component may be approximated by a collection of 2-D sinusoids which give rise to 2-D delta functions in the spectral domain. Thus, each evanescent component (whose "spectral density" is a 1-D delta function) is approximated for the purpose of this test by a linear combination of harmonic components (whose "spectral densities" are 2-D delta functions).

In order to estimate the deterministic components, we first choose a threshold and admit all frequencies where the periodogram is above the threshold as deterministic component frequencies. The threshold must be carefully chosen such that, (a) we extract only the sharp peaks in the periodogram, and (b) the number of harmonics obtained is a small fraction of the total number of DFT bins. The latter condition is required because, theoretically, the support of the deterministic component is of Lebesgue measure zero in the spectral domain. A good choice of threshold can be obtained by plotting the threshold against the number of deterministic component frequencies given by that threshold. It is shown in [17] that when a texture has a deterministic component the plot exhibits a relatively flat region and a sharp "knee" near zero, indicating the presence of sharp peaks in the periodogram. The threshold is then chosen in the relatively flat portion close to the knee in the curve. On the other hand, the plot for a purely-indeterministic (random looking) texture has a slow decay indicating the absence of deterministic components. We refer the interested reader to [17] for additional details.

In case existence of a deterministic component is established, a second test is applied to distinguish the harmonic from the evanescent components. The proposed test employs a modified version of the Hough transform for detecting lines in 2-D arrays. This modification is based on the *a-priori* knowledge that the spectral support of each one of the evanescent components is a line in the frequency plane such that its slope is defined by two coprime integers α and β. In the presence of evanescent components, the periodogram peaks identified by the previous test, are concentrated along lines. On a finite dimension observed field, only a finite number of (α, β) pairs may be defined. (This is since α and β are integers representing distances between consecutive samples along the "rows" and "columns" defined with respect to the RNSHP total-ordering definition $(\alpha, \beta) \in O$). Therefore, for given dimensions of the observed field, we search among all possible combinations of the spectral support parameters α and β for the pairs of (α, β) that best explain the concentration of peaks along lines in the frequency plane. If at least one such line is found the deterministic field is considered to be the sum of evanescent components. Otherwise, we conclude that the spectral peaks are isolated (*i.e.*, they are not aligned along lines) and hence they are due to harmonic components.

Following this decision making step, we turn to the actual estimation of the components' parametric models. In case the observed texture was found by the foregoing two tests to contain harmonic components, we employ an iterative procedure to estimate the texture parameters. In each iteration the frequency of the dominant harmonic component is estimated by

$$(\hat{\omega}_p, \hat{\nu}_p) = \underset{(\omega,\nu)}{\text{argmax}} \left| \text{DFT}\Big(y(n,m) \Big) \right| . \tag{6}$$

Multiplying the observed signal $y(n,m)$ by $\cos(\hat{\omega}_p, \hat{\nu}_p)$ and evaluating the mean of this signal we obtain an estimate of the amplitude coefficient C_p. This is since the mean of all the cosinusoidal and sinusoidal terms of $y(n,m)\cos(\hat{\omega}_p, \hat{\nu}_p)$ tends to zero while the DC term tends to C_p. Thus

$$\hat{C}_p = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} y(n,m) \cos(\hat{\omega}_p, \hat{\nu}_p) \tag{7}$$

and similarly

$$\hat{D}_p = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} y(n,m) \sin(\hat{\omega}_p, \hat{\nu}_p) \tag{8}$$

where the dimensions of the observed image are $N \times M$. Next we subtract the estimated harmonic component from the observed signal $y(n,m)$ and repeat this procedure iteratively until all harmonic components whose magnitude is higher than the foregoing test threshold are extracted. The residual field is the purely-indeterministic component of the texture.

In case the observed texture was found by the foregoing two tests to contain evanescent components, we employ the estimation procedure derived in [20] to estimate their parameters. This estimation procedure employs the properties of the covariance matrix of the field deterministic component. Let us denote the number of evanescent components of the field by Q. It is then shown in [20] that in the absence of a purely-indeterministic component

$$\mathrm{rank}(\Gamma_d) = S\sum_{k=1}^{Q}(2-\delta(\nu_k))|\alpha_k| + T\sum_{k=1}^{Q}(2-\delta(\nu_k))|\beta_k| - \sum_{k=1}^{Q}(2-\delta(\nu_k))|\alpha_k|\sum_{k=1}^{Q}(2-\delta(\nu_k))|\beta_k| \tag{9}$$

where $\delta(\nu)$ denotes the Dirac delta, and Γ_d is the covariance matrix of the deterministic field.

In the presence of the purely-indeterministic component the covariance matrix of the field is given, due to the orthogonality of the fields, by $\Gamma = \Gamma_d + \Gamma_{PI}$, where Γ_{PI} denotes the covariance matrix of the purely-indeterministic field. In the presence of the purely-indeterministic component Γ is a full rank matrix, yet its dominant eigenvalues are due to the evanescent components. It is therefore concluded that by determining the number of dominant eigenvalues of the covariance matrix for various subfields of the observed field, we obtain a system of equations, such that its solutions provide us the set of possible values the parameters (α, β) can assume. Note from the evanescent field model (4) that for a fixed $c = n\alpha - m\beta$ (*i.e.*, along a line on the sampling grid), the samples of the evanescent component are nothing but the samples of constant amplitude harmonic signal, whose frequency is $\frac{\nu_i^{(\alpha,\beta)}}{\alpha^2+\beta^2}$. Thus, having estimated the (α, β) pairs, the corresponding frequency parameter of each evanescent component can be easily evaluated using standard techniques (such as locating the maxima of the absolute value of this 1-D signal DFT).

Once the spectral support parameters of each evanescent component, $(\alpha, \beta), \nu_i^{(\alpha,\beta)}$, were estimated, it can be demodulated in a similar way to the one applied for estimating the amplitude parameters of the harmonic components. However, in this case the averaging is performed along a line on the sampling grid such that $c = n\alpha - m\beta$. Thus

$$\hat{s}_i^{(\alpha,\beta)}(c) = \frac{1}{N_s}\sum_{n\hat{\alpha}-m\hat{\beta}=c} y(n,m)\cos\left(2\pi\frac{\hat{\nu}_i^{(\alpha,\beta)}}{\hat{\alpha}^2+\hat{\beta}^2}(n\hat{\beta}+m\hat{\alpha})\right) \tag{10}$$

and similarly

$$\hat{t}_i^{(\alpha,\beta)}(c) = \frac{1}{N_s}\sum_{n\hat{\alpha}-m\hat{\beta}=c} y(n,m)\sin\left(2\pi\frac{\hat{\nu}_i^{(\alpha,\beta)}}{\hat{\alpha}^2+\hat{\beta}^2}(n\hat{\beta}+m\hat{\alpha})\right) \tag{11}$$

where N_s denotes the number of the observed field samples that satisfy the relation $n\alpha - m\beta = c$. This procedure provides estimates of the 1-D sequences $\{s_i^{(\alpha,\beta)}(n\alpha - m\beta)\}$ and

$\{t_i^{(\alpha,\beta)}(n\alpha - m\beta)\}$ of each evanescent field. Finally, these sequences are fitted with their 1-D AR models. The removal of all evanescent components of the field leaves us with a residual field which is the purely-indeterministic component of the texture.

The parameters of the purely-indeterministic component are estimated using a computationally efficient algorithm for estimating its moving average model, [21]. The algorithm first fits a 2-D NSHP AR model to the observed field, by using the ML algorithm, [15]. Note that in this case, where all the deterministic components have already been removed, the procedure of obtaining a maximum-likelihood estimate of the AR model parameters is reduced to a solution of a linear least squares problem. In the second stage, the estimated parameters of the AR model are employed to compute the parameters of the moving average model, through a least squares solution of a system of linear equations. We refer the interested reader to [21] for a detailed derivation and analysis of the algorithm and its statistical performance.

4 Distance Measures for Textures

It is shown in [16] that assuming the texture field is a realization of a Gaussian random field with mixed spectral distribution, essentially indistinguishable replicas of the original texture are synthesized from the estimated parameters. We therefore adopt the Gaussian assumption for the derivation of a distance measure, as well as for the image segmentation application. In this framework it is assumed that each observed texture is a realization of a Gaussian random field whose mean is related to the harmonic component of the Wold decomposition and the structure of its covariance matrix is determined by the parameters of the evanescent and purely indeterministic components.

4.1 The Kullback Distance

Having adopted the Wold decomposition based random field model as our texture model we face the problem of formulating a definition for a distance measure between any two given fields. In other words, having estimated the parametric model (1)-(3) of each field, how can one test whether these two fields are different samples of the same texture, or how close these two textures are one to the other, etc. Since the different components of the texture model have different stochastic models it is clear that the required distance measure *must* incorporate in a natural way the distances of the different components into a single measure. By doing so, the heuristic question of the determination of the relative weighting of the distances between the different components of the model, is avoided. (We note here that an earlier attempt to employ the same texture model for indexing images was made in [11], where an *ad-hoc* procedure was developed in order to incorporate the distances between the purely-indeterministic components and the harmonic components into a single distance measure.) It should be emphasized that in the context of this paper the term distance refers to a measure of how far away from each other are the probability distributions of the observed texture fields, and it should *not* be interpreted in the strict sense it has in metric spaces.

More specifically, in this work we have chosen the Kullback distance [18] between two Gaussian models of textures.

$$D(P_1, P2) = \frac{1}{2}(\mu_2 - \mu_1)^T (\sum_1^{-1} + \sum_2^{-1})(\mu_2 - \mu_1) + \frac{1}{2}\mathrm{tr}(\sum_1^{-1} \sum_2 + \sum_1 \sum_2^{-1} - 2I) \quad (12)$$

This specific distance measure was chosen due to its generality and its ability to incorporate all the model components into a single distance measure. We note here that an alternative choice is provided by the Bhattacharyya distance [18]. The Kullback distance was preferred, in this case due to its lower computational requirements.

Clearly, one is interested in being able to define a distance measure, which is invariant to translation and rotation. Since the observed textures are homogeneous, translation affects only the mean function, but not the covariance. Rotation affects both the mean and covariance. To achieve the desired invariance, we have implemented the following procedure:

- Estimate the parametric model of each texture patch introduced to the system using the algorithm derived in Section 3.
- For textures with harmonic components, the texture is first rotated so that its dominant harmonic component is aligned with a predefined direction (the x-axis, for example).
- For textures with evanescent components, the dominant evanescent component is considered the one whose modulating 1-D purely-indeterministic processes have maximal variance. Thus, the texture is first rotated so that its dominant evanescent component has $(\alpha, \beta) = (1, 0)$.
- For textures with harmonic components where the phase of the dominant component is not zero, we crop a sub-picture of the original in which the phase of the dominant component is zero.
- Re-estimate the texture parameters to find the parameters of all the model components.

Having estimated the (normalized) parametric models of the two textures whose distance is to be evaluated, one has to compute the mean vector and covariance matrix of each field. The Kullback distance between these two fields can then be evaluated by substituting in (12) the mean vector and covariance matrix of each field with their estimates. Evaluation of the mean vector is performed by synthesizing the harmonic field from its estimated parametric model, followed by vectorization of the 2-D field into a vector form.

Since the purely-indeterministic component and all the evanescent components are mutually orthogonal the estimated covariance matrix of each field is obtained by summing the covariance matrices of its evanescent and purely-indeterministic components. An expression for the covariance matrix of an evanescent field in terms of the field parameters is derived in [19]. In this paper we obtain an estimate of the covariance matrix of each evanescent component by substituting the unknown field parameters by their estimates. The procedure for computing the covariance matrix of the purely-indeterministic component from its MA model is detailed in [21]. We note that this derivation of the estimated covariance matrix *guarantees* that it is positive semi-definite.

4.1.1 Experimental Results

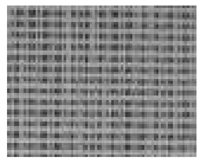
To test the distance measures, we chose a reference texture, and evaluated its distance from other texture patches. To perform the tests we constructed a database of Brodatz textures. The database is formed by cropping, shifting and rotating (by 30, 60, and 90 degrees) more than 30 Brodatz textures. The total number of textures in the database is over 500. The goal of the experiment was to arrange all the patches in the database in an ascending order with respect to their distance from the given test texture (*i.e.*, the database texture that is the "nearest" to the test texture should be ranked first, etc.) The experimental results of two such tests are depicted in Figure 1 and Figure 2. The textures are ordered in an ascending order from left to right and from top to bottom. Note that, as required, the rotated versions of the test texture were found to have the minimal distance to that texture, among all the database texture. Ordered next are textures that have structure close to that of the test texture.

Thus, the Kullback distance provides a unified and rigorous solution to the problem of grouping together the distances of the different stochastic models of the Wold decomposition components into a single measure. However, the procedure of computing this distance is expensive in terms of both computational complexity and memory requirements due to the need to invert the covariance matrix of each field. (The covariance matrix of a field whose dimensions are $N \times M$ is $MN \times MN$). We therefore investigate in the next section a hierarchical evaluation of the Kullback distance that allows us to save the computation of the fields' covariance matrices and their inverses.

4.2 A Computationally Efficient Approach for Evaluating the Distance Between Textures

In this section we present an alternative method for computing the Kullback distance such that the direct evaluation of the fields' covariance matrices and their inverses, is avoided. This method provides a computationally efficient approximation of the distance. Hence, it is more suitable for database retrieval applications. In the proposed method computation of the distance is implemented in a *hierarchical* way, rather in the joint approach of the Kullback distance evaluation.

This hierarchical approach is directed by the estimated parametric models of the two textures whose distance is to be evaluated. If one texture is found to have components that the other does not have, the distance between these two is set to infinity. For example, if one texture has a deterministic component while the other is found to be purely-indeterministic, it is clear that the textures are different. Hence, the distance is set to infinity.



Test Texture

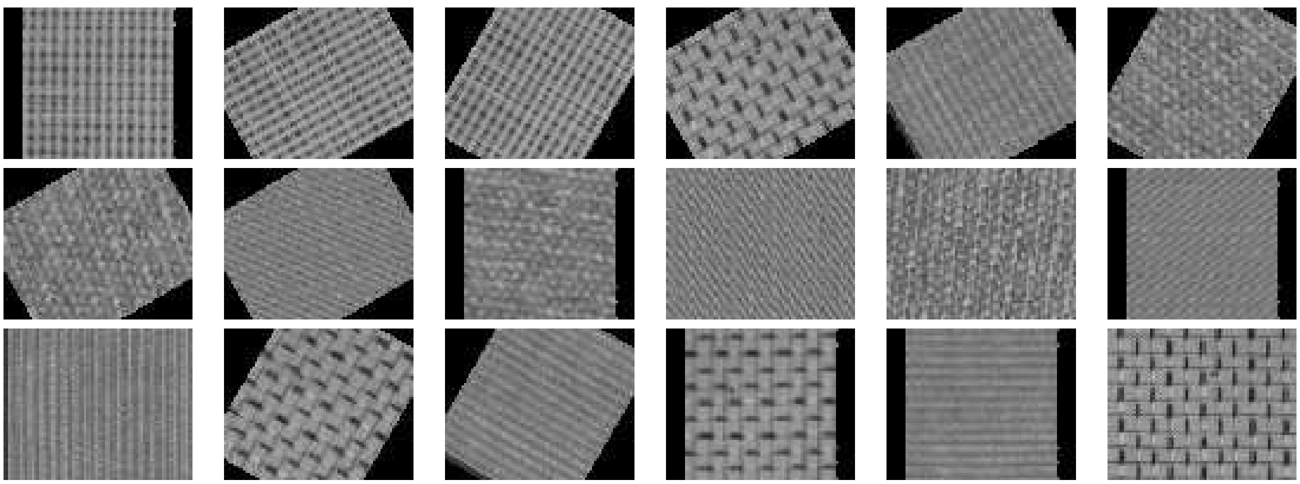


Figure 1: Textures ordered using the Kullback distance.

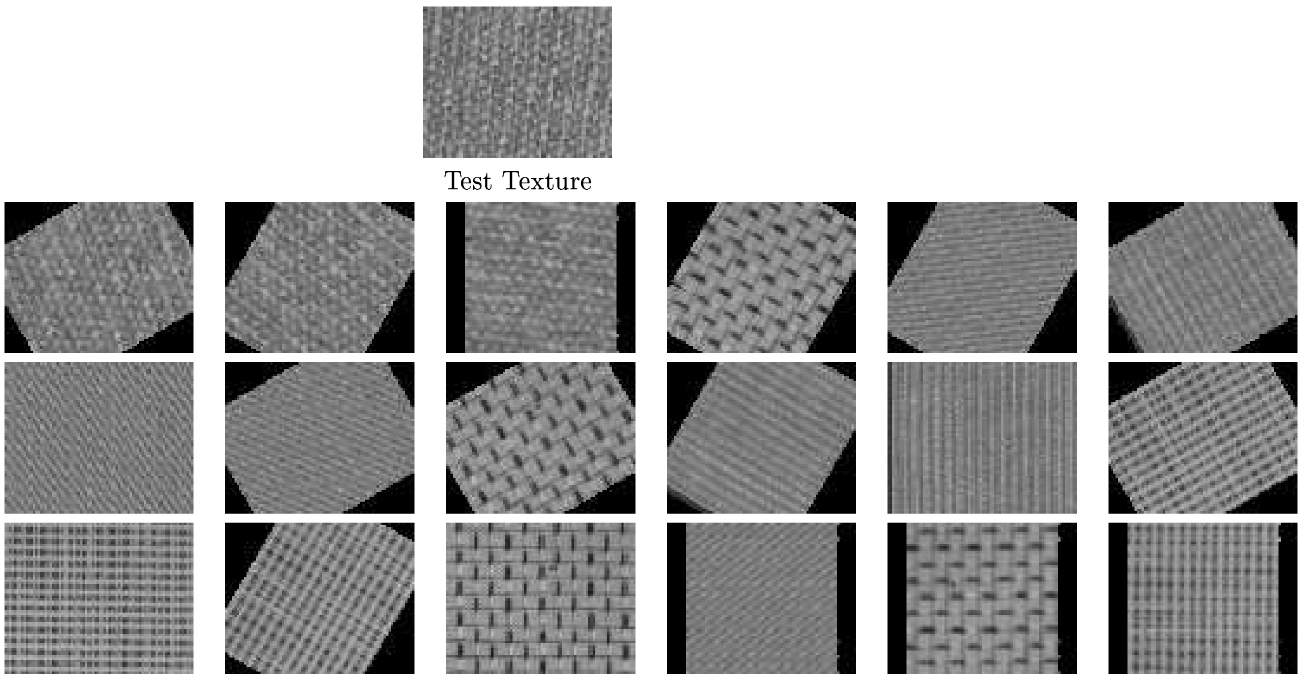


Figure 2: Textures ordered using the Kullback distance.

In case both textures have only a purely-indeterministic component, the Kullback distance is reduced to the Itakura-Saito distance, [18],

$$d_{IS}(s_1, s_2) = \left\| \frac{s_1}{s_2} - \log \frac{s_1}{s_2} \right\|_2 \tag{13}$$

where s_1 and s_2 denote the spectral densities of the two fields. Hence, in the case where both textures are dominated by their purely-indeterministic components the Kullback distance can be approximated by the Itakura-Saito distance between the spectral densities of the MA modeled purely-indeterministic components.

Next, we have from (12) that when both purely-indeterministic components are white noise fields with the same variance, and no evanescent components exist in both fields, the Kullback distance is reduced to the Euclidean norm of the means difference. Thus, in case both textures have dominant harmonic components, *i.e.*, the contribution of the purely-indeterministic components can be neglected for the purpose of distance evaluation, we might as well assume that both purely-indeterministic components are white noise fields with the same (small) variance. Thus, the distance between the two textures is computed as the Euclidean distance between the magnitudes of the Fourier transforms of the harmonic components of these textures.

In the case where both textures have dominant evanescent components an evaluation of (12) is still required.

4.2.1 Experimental Results

To test the hierarchical distance evaluation, we tested its performance using the same procedure and the same database described above. To illustrate the performance of the suggested procedure, different types of test textures were chosen in the next examples. The test textures in figures 3-5 are structured and dominantly harmonic, while in Figure 6 the test texture is dominantly evanescent, and in Figure 7 the test texture is purely-indeterministic. We therefore conclude that the proposed hierarchical approach for approximating the Kullback distance between texture fields provides the desired results (with respect to a human viewer decision), at a much lower computational complexity than the direct evaluation of (12).

5 Unsupervised Image Segmentation

Segmentation is an essential and critical step in the process of labeling any given image according to its textural components. The aim of this step is to identify the different texture areas in any given image, to estimate their boundaries and subsequently their parameters. By doing so each textured region of the image is characterized by its estimated parametric model. The set of different texture models that correspond to different regions of a given image, form the set of indices (labels) of that image.

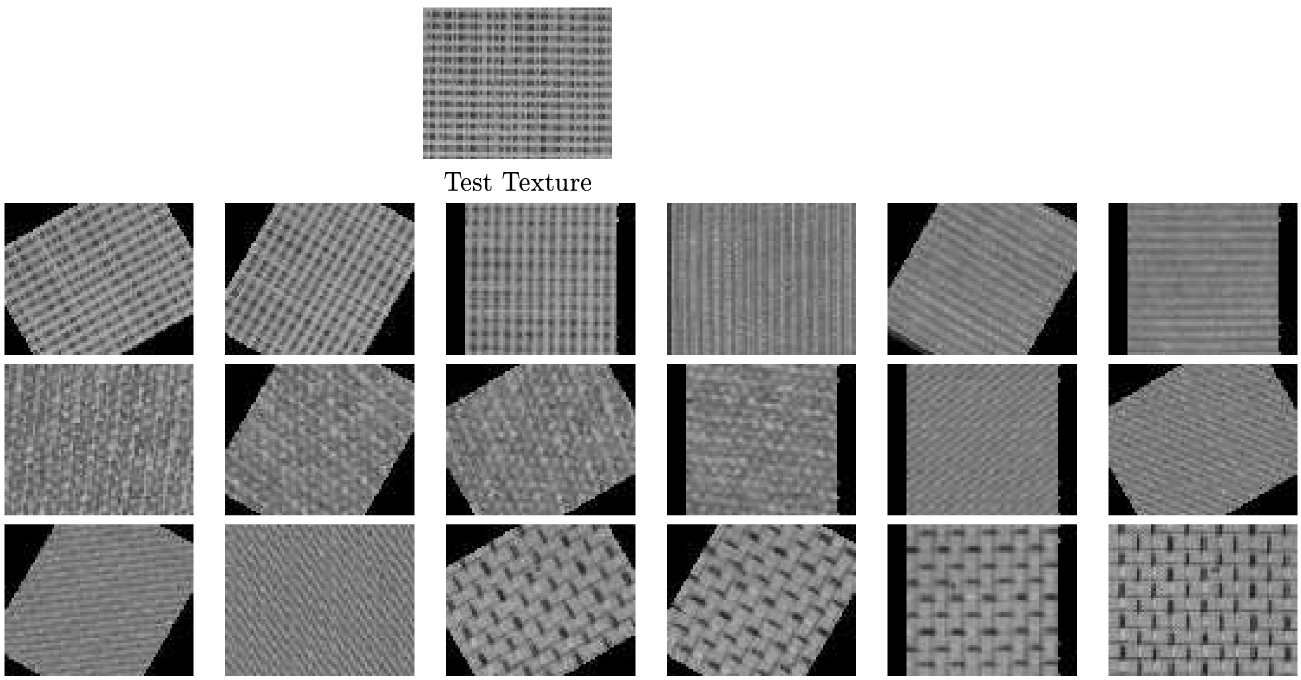


Figure 3: Textures ordered using the hierarchical distance. The test texture is structured, dominantly harmonic.

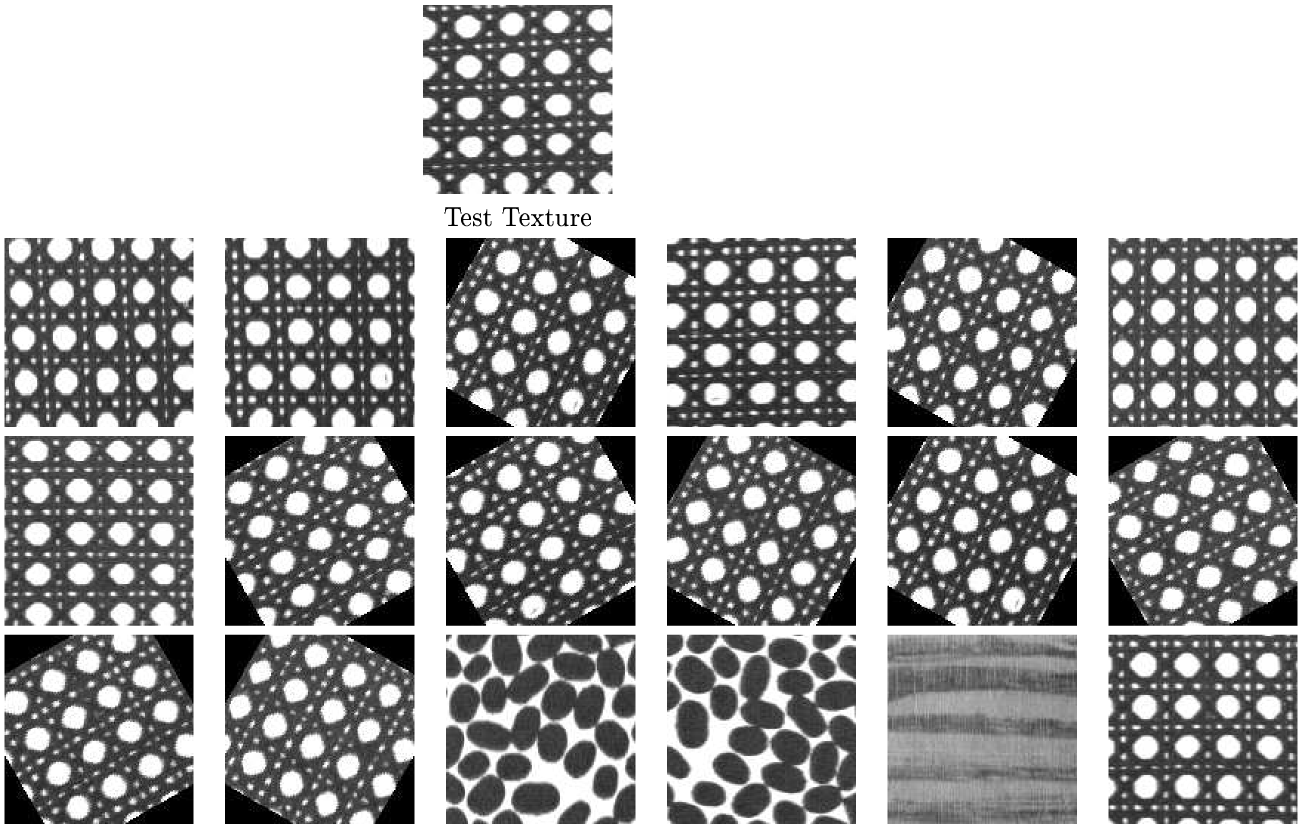


Figure 4: Textures ordered using the hierarchical distance. The test texture is structured, dominantly harmonic.

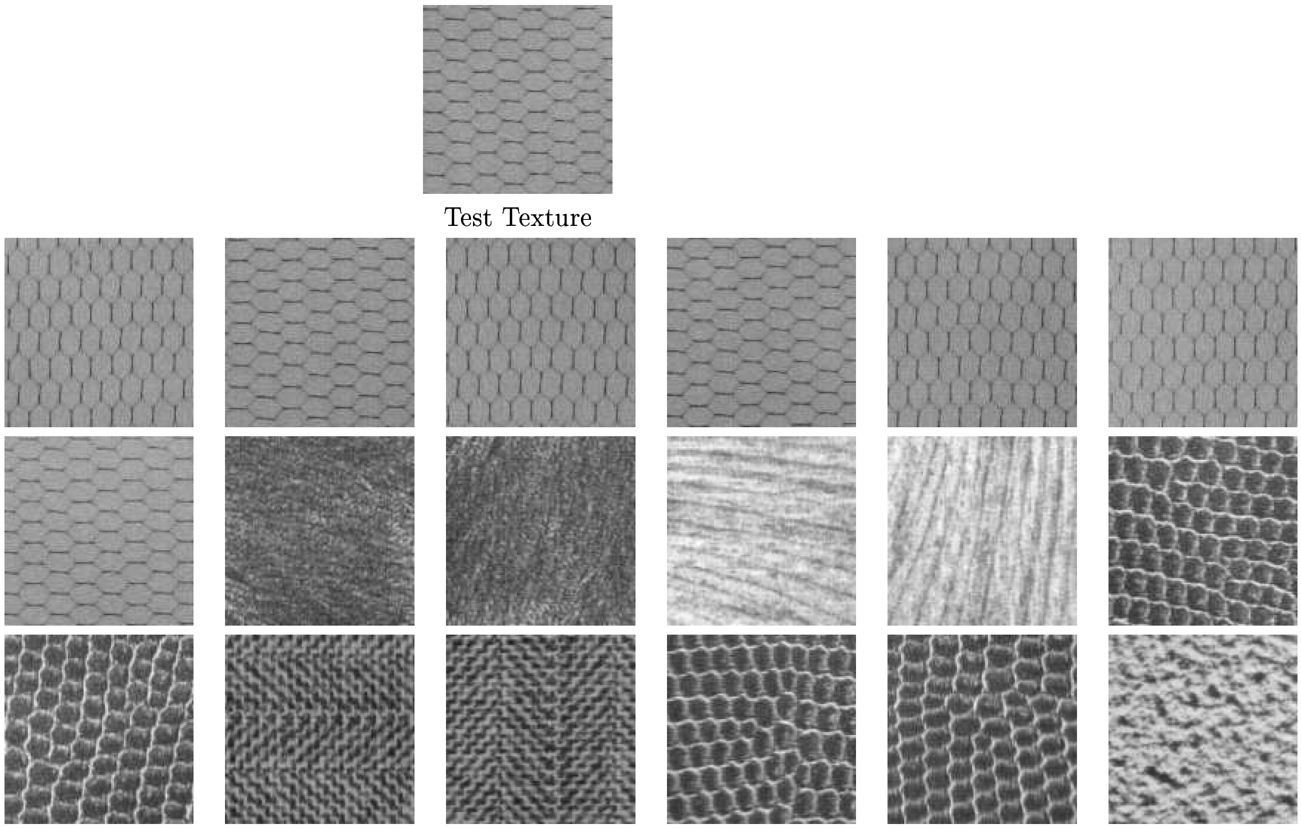
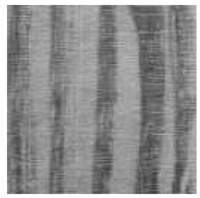


Figure 5: Textures ordered using the hierarchical distance. The test texture is structured, dominantly harmonic.



Test Texture

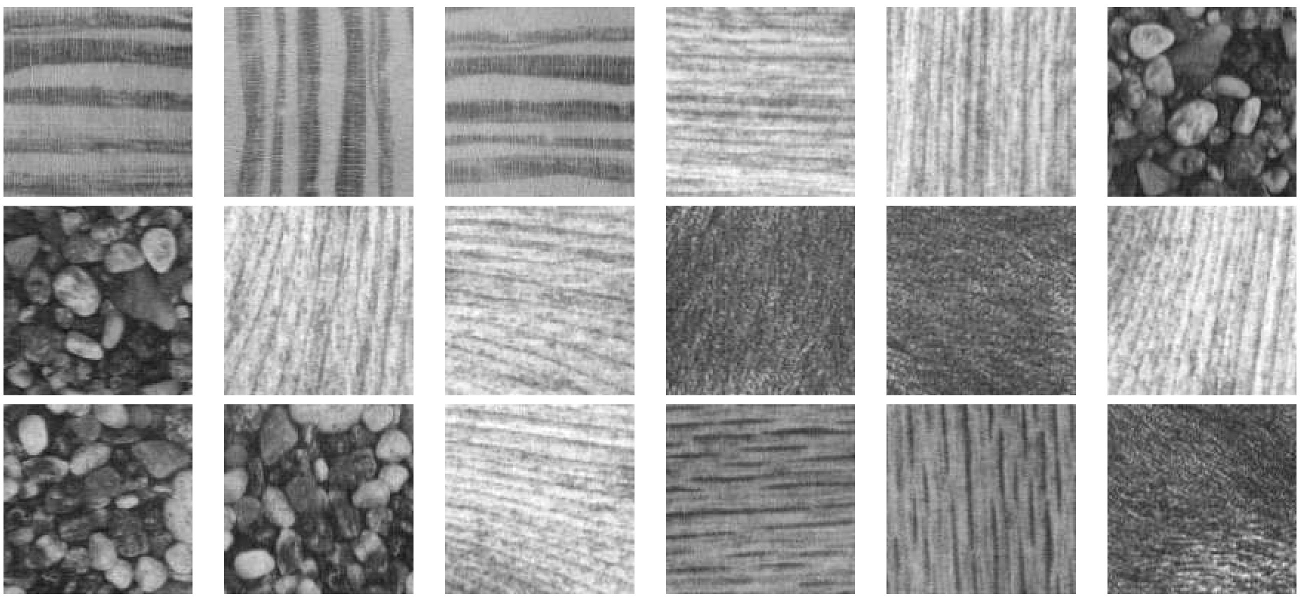
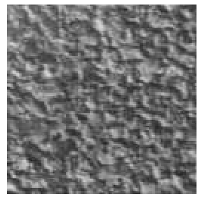


Figure 6: Textures ordered using the hierarchical distance. The test texture is structured, dominantly evanescent.



Test Texture

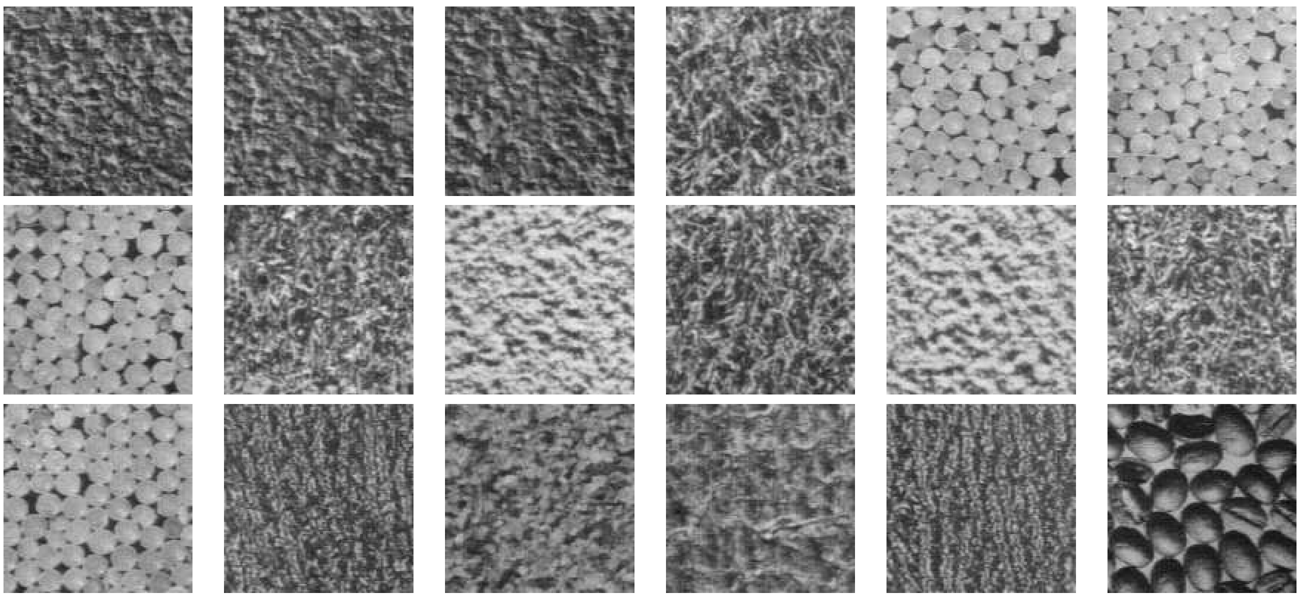


Figure 7: Textures ordered using the hierarchical distance. The test texture is purely-indeterministic.

In this section we use the Wold decomposition based texture model as a tool for segmenting the image into its distinct textured regions. The proposed unsupervised segmentation procedure models the image using a doubly stochastic Gaussian (DSG) model [7]. A Markov random field (MRF) model is applied to model the label field (which indicates the texture to which each pixel belongs) and the 2-D Wold decomposition based model (1)-(5) is used to model each of the textures (the texture model parameters are conditioned on the label field).

The difficulty of the unsupervised segmentation problem is due to the fact that statistically reliable estimation of the parameters of the deterministic component of a texture patch requires the patch to be large enough so that at least few periods of each harmonic component are available. On the other hand, it is obvious that arbitrary division of the image into blocks may put several textures into one block and thus lead to useless estimation results. Hence the proposed segmentation procedure is performed by a two step procedure. An initial coarse parameter estimate is obtained by dividing the image into small blocks, fitting a low order NSHP AR model to each block, and clustering the estimated parameters. The clustering is performed in the AR model parameter space using the K-means algorithm. The distance used to cluster the blocks is the Itakura-Saito distance of the estimated spectral densities (13).

Having obtained an initial segmentation, the texture model parameters can be estimated from the larger patches that result from this initialization step. This segmentation result is then iteratively refined in the next stage using a Markovian prior for the labels field, where in each iteration the texture model parameters are re-estimated for each patch.

More specifically, the estimated parametric models of the individual textures are employed to construct an energy function. This energy function is used in a simulated annealing algorithm to obtain maximum *a-posteriori* (MAP) estimates of the label field. The label field estimation is iteratively performed until the algorithm converges, *i.e.*, until the number of label changes in one iteration falls below a preset threshold. The optimal estimate of the label field at site s is given by:

$$f^* = \min_{f \in F}\{U_s(f|d)\} = \min_{f \in F}\{U_s(d|f) + U_s(f)\} \tag{14}$$

where f is the label and d is the data field. The energy $U_s(d|f)$ is evaluated at each site s using the following procedure:

- Let $W(s)$ be some window centered at s. Let $F^d(s)$ be the magnitude of the DFT of the observed image, evaluated in that window.
- Using the texture model parameters that correspond to the current label f of the site, we synthesize the texture, and evaluate the magnitude of the synthesized field DFT, $F^f(s)$, in the same window $W(s)$.
- The conditional energy is then given by:

$$U_s(d|f) = \|F^d(s) - F^f(s)\|_2 \tag{15}$$

Note that by adopting (15) as our estimate of the conditional energy, phase errors due to possible translation discrepancies, are eliminated. Let C be the set of cliques corresponding to some neighborhood system. Here the field is assumed to be first order Markov. The regularization energy at site s is then defined using a Potts model:

$$U_s(f) = -K \sum_{<r,s>\in C} \delta(f_r, f_s) \tag{16}$$

Figures 8-10 depict the results of applying the proposed segmentation algorithm to example images. In all the examples the final segmentation result is obtained after 100 iterations of the algorithm.

Finally, in the foregoing discussion it was assumed that the number of clusters K, is *a priori* known. Unfortunately in practice, this is not the case. Hence in the proposed algorithm K is over estimated and we assume an arbitrary (large) number for K. The example results demonstrate that using simulated annealing, the procedure converges to a reasonable number of clusters.

6 Conclusions

We have presented a parametric method for indexing and retrieval of multimedia data. The proposed indexing and retrieval strategy is based on the usage of textural information contained in the data imagery components as the indexing keys. The parametric texture model is applied both to segment each image into distinct texture regions, as well as to derive a distance measure that rigorously incorporates the distances between the different stochastic models of the textures' components into a single distance measure. Since the same parametric texture model is used for coding, storing, and indexing the texture data, the proposed parametric indexing scheme provides a representation that facilitates content-based retrieval of the compressed data. It is shown that the proposed segmentation, distance evaluation, and indexing methods produce comparable results to those obtained by a human viewer.

7 Acknowledgments

The authors wish to thank Prof. Badri Roysam for allowing them to use the image in Fig. 8a.

References

[1] R. Stoica, "Etude de la decomposition de Wold 2D et son application aux images texturees", DEA Thesis, ENSEEIHT, Toulouse University, Sept. 1997

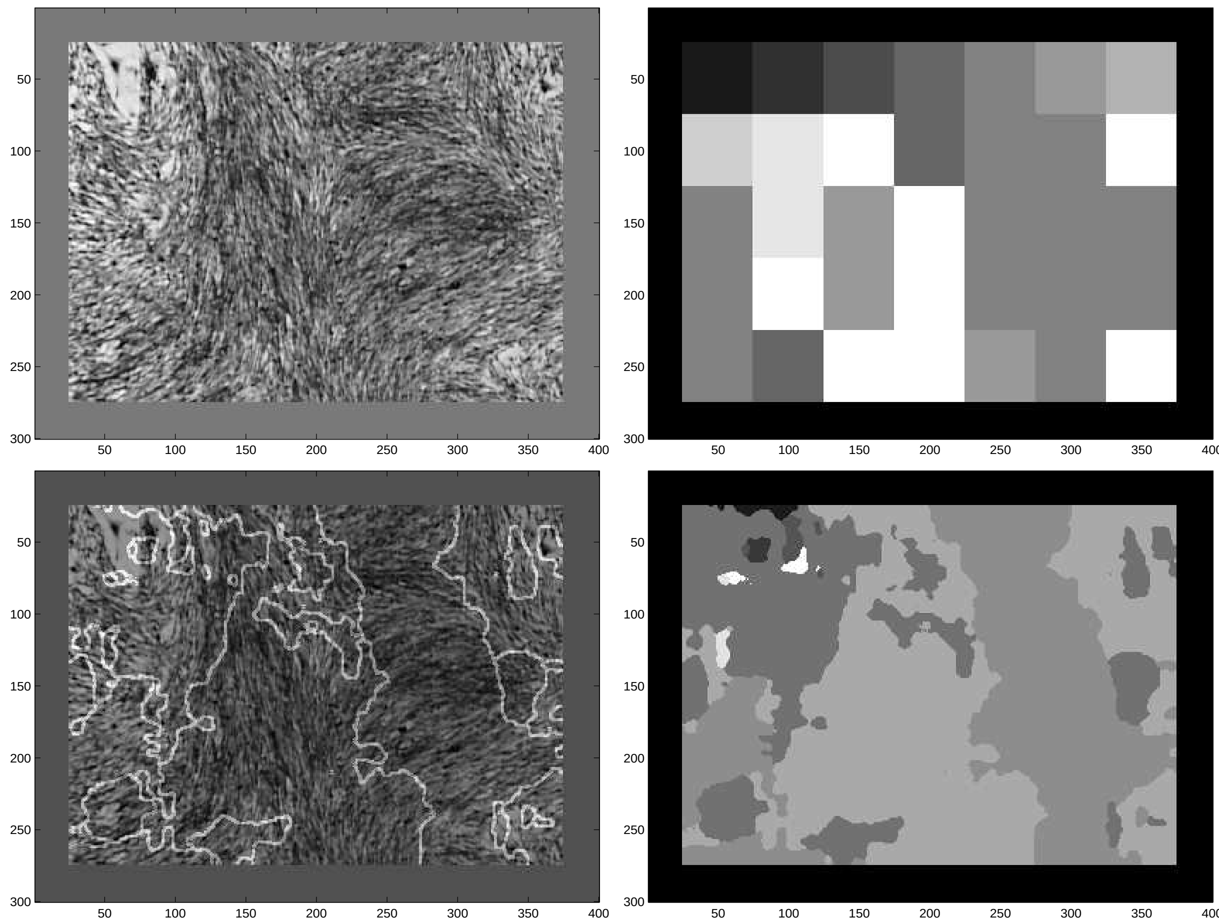


Figure 8: Unsupervised segmentation of a medical image: Clockwise from top-left: Original image; Coarse segmentation map; Final segmentation map; Final segmentation result.

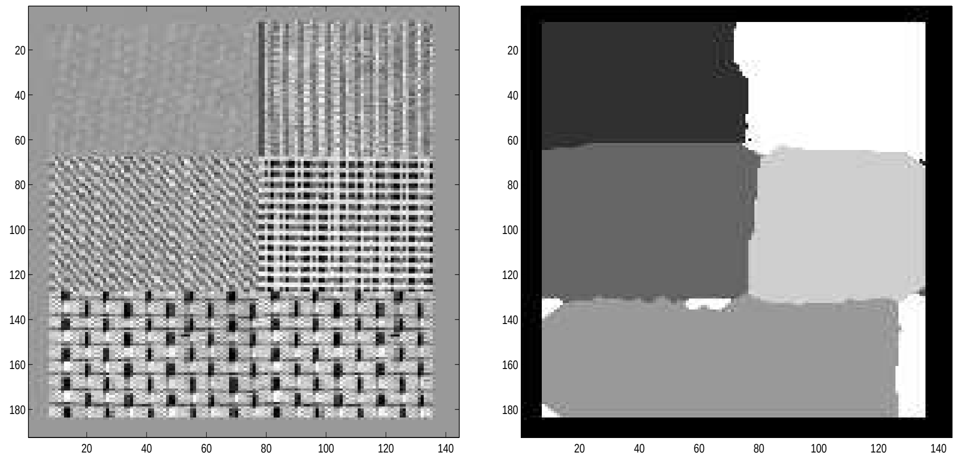


Figure 9: Unsupervised segmentation of Brodatz textures: Original image (left); Segmentation with the SA algorithm after 100 iterations (right).

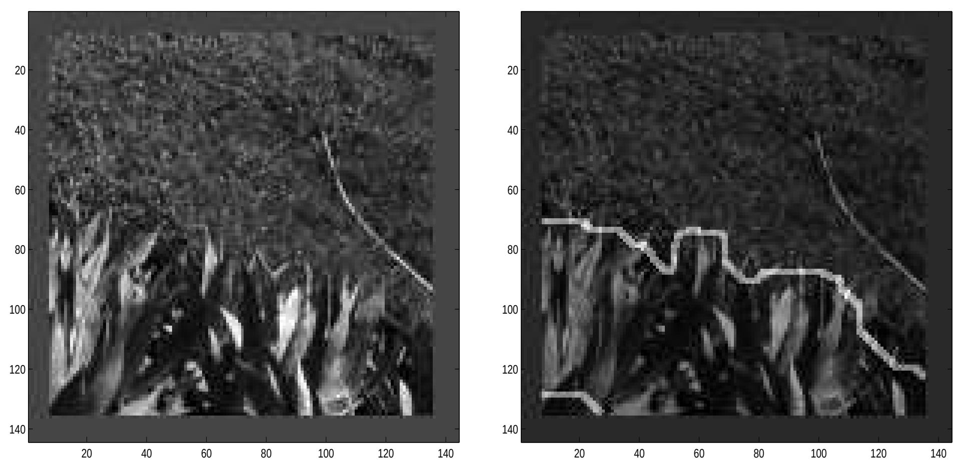


Figure 10: Unsupervised segmentation of a natural scene: a)Original image (left); Segmentation with the SA algorithm after 100 iterations (right).

[2] R. Stoica, J. Zerubia, and J. M. Francos, "The Two-Dimensional Wold Decomposition for Segmentation and Indexing in Image Libraries", *IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 5, pp. 2977-2980, Seattle, 1998.

[3] R. Stoica, J. Zerubia, and J. M. Francos, "Image Retrieval and Indexing: A Hierarchical Approach in Computing the Distance Between Textured Images", *IEEE Int. Conf. Image Processing*, vol. 2, pp. 794-799, Chicago, 1998

[4] G. L. Lohse and A. R. Rao, "Towards a Texture Naming System: Identifying Relevant Dimensions of Texture" *Proc. IEEE Conf. on Visualization*, pp. 220-227, San Jose, CA, 1993.

[5] J. Mao and A. K. Jain "Texture Classification and Segmentation Using Multiresolution Simultaneous Autoregressive Models" *Pattern Recognition*, vol. 25, No. 2, pp. 173-188,1992.

[6] C. Bouman and B. Liu, "Multiple Resolution Segmentation of Textured Images,"*IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 99-113, Feb. 1991

[7] F. C. Jeng and J. W. Woods, "Simulated Annealing in Compound Gaussian Random Fields,"*IEEE Trans. on Information Theory*, vol. 36, pp. 94-107, Jan. 1990

[8] T. Chang and C.-C. J. Kuo, "Texture Analysis and Classification with Tree-Structured Wavelet Transform", *IEEE Trans. on Image Processing*, vol. 2, pp. 429-441, 1993.

[9] M. Unser, "Texture Classification and Segmentation Using Wavelet Frames", *IEEE Trans. on Image Processing*, vol. 4, pp. 1549-1560, 1995.

[10] R. Mohr, P. Gros, B. Lamiroy, S. Picard, and C. Schmid, "Indexation et recherche d'images," *Proc. GRETSI*, Grenoble, Sept. 15-19, 1997.

[11] F. Liu and R. Picard, "Periodicity, Directionality and Randomness: Wold Features for Image Modeling and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 722-733, 1996.

[12] Y. Wu and Y. Yoshida, "An Efficient Method for Rotation and Scaling Invariant Texture Classification", *IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 2519-2522, 1995.

[13] R. Sriram, J. M. Francos, and W. A. Pearlman, "Texture Coding Using a Wold Decomposition Based Model," *IEEE Trans. Image Process.*, vol. 5, pp. 1382-1386, 1996.

[14] J. M. Francos, A. Z. Meiri and B. Porat, "A Wold-Like Decomposition of 2-D Discrete Homogeneous Random Fields," *Annls. Appl. Prob.*, vol. 5, pp. 248-260, 1995.

[15] J. M. Francos, A. Narashimhan, and J. W. Woods, “Maximum-Likelihood Parameter Estimation of Discrete Homogeneous Random Fields with Mixed Spectral Distribution,” *IEEE Trans. Signal Process.*, vol. 44, pp. 1242-1255, 1996.

[16] J. M. Francos, A. Narashimhan, and J. W. Woods, “Maximum-Likelihood Estimation of Textures Using a Wold Decomposition Model”, *IEEE Trans. Image Process.*, vol. 4, pp. 1655-1666, 1995.

[17] R. Krishnamurty, J. W. Woods, and J. M. Francos, “Adaptive Restoration of Textured Images with Mixed Spectra”, *IEEE Trans. Image Process.*, vol. 5, pp. 648-652, 1996.

[18] M. Basseville, “Distance Measures for Signal Processing and Pattern Recognition,”*Signal Processing*, vol. 18, pp. 349-369, 1989.

[19] J. M. Francos, “Bounds on the Accuracy of Estimating the Parameters of Discrete Homogeneous Random Fields with Mixed Spectral Distributions," *IEEE Trans. Info. Theory.*, vol. 43, pp. 908-922, 1997.

[20] G. Cohen and J. M. Francos, “Efficient Parameter Estimation of Evanescent Random Fields," submitted for publication.

[21] J. M. Francos and B. Friedlander, “Parameter Estimation of Two-Dimensional Moving Average Random Fields," *IEEE Trans. Signal Process.*, vol. 46, pp. 2157-2165, 1998.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399