A Wavelet-Based Approach for Analyzing Industrial Stochastic Textures With Applications

Jacob Scharcanski, *Senior Member, IEEE*

Abstract—**Several continuous manufacturing processes use stochastic texture images for quality control and monitoring. Large amounts of pictorial data are acquired, providing important information about both the materials produced and the manufacturing processes involved. However, it is often difficult to measure objectively the similarity among industrial stochastic images or to discriminate between texture images of stochastic materials with distinct properties. Nowadays, the degree of discrimination required by industrial processes often goes beyond the limits of human visual perception. This paper proposes to model this specific class of textures as colored noise and presents a new approach for multiresolution stochastic texture representation and discrimination in industry (e.g., nonwoven textiles and paper). The wavelet transform is used to represent stochastic texture images in multiple resolutions and to describe them using local orientation and density variability as features. Based on this representation, a multiresolution distance measure for stochastic textures is proposed, and industrial applications of the method and experimental results are reported. The conclusions include ideas for future work.**

Index Terms—**Anisotropy, colored noise, industrial quality control, maintenance, nonwoven textiles, stochastic textures, wavelets.**

I. INTRODUCTION

MODERN computer-based control systems are able to collect large amounts of information in industry, display it to operators, and store it [1]. However, decisions are often made by operators on an *ad hoc* basis with little computer support, whereas the stored process data could also be used in tasks such as 1) automatic analysis and interpretation of process operational data (either in real time or over the operational history); 2) development of intelligent state-based methods for process monitoring, control, and diagnosis; and 3) data mining and knowledge discovery in manufacturing, for purposes such as maintenance and training.

In several industrial continuous processes, static and dynamic stochastic texture images are acquired and used in quality control. In particular, stochastic texture images are used in manufacture of foil-like materials such as nonwoven textiles, paper, polymer membranes, conductor, and semiconductor coatings. Important information can be extracted from these gray-level images, representing spatial density variations of such materials. In general, these images contain a composition of stochastic features at various scales, resulting from local clumping and aligning of constituent matter with varying degree of regularity.

Possible benefits from the analysis of such stochastic textures include: 1) interpretation of temporal process operational data based on texture formation anisotropy; 2) system diagnosis and monitoring by accessing the image collection in terms of local spatial density variation, as well as other textural features; and 3) *a posteriori* data analysis to reveal influential factors. For example, the early detection of patterns in temporal and/or spatial variability is a common problem in nonwoven textiles manufacturing, which can be approached by multiresolution stochastic texture analysis, as we will detail later.

Often, industrial machine operators try to evaluate stochastic texture images visually and estimate the manufacturing process condition using their experience in the field. This empirical approach is subjective and prone to failure, mainly because human vision is limited in terms of its ability to distinguish between stochastic textures [2]–[4]. Despite advances in texture representation and classification over the past three decades [3], [5]–[7], the problem of stochastic texture feature interpretation and classification remains a challenge for researchers [8] and for a large segment of industry [2]; this is mostly because in several practical situations, stochastic texture analysis and classification must outperform visual texture discrimination and match industrial needs.

A variety of methods has been proposed for extracting texture features from textured images, e.g., geometric, random field, fractal, and signal processing models for textures. Most recent works on textures tend to concentrate on two areas, namely: 1) multichannel filtering theory and 2) statistical modeling [5]. A substantial part of the work on multichannel filtering theory is inspired by neurophysiological principles [9]. On the other hand, statistical modeling focuses on characterizing textures as probability distributions and using statistical theory to formulate and solve texture processing problems mathematically. Wavelet-based texture characterization can integrate the above two aspects and has attracted attention recently because of its usefulness in several important applications, such as texture classification [6], [10] and texture segmentation [11]. These approaches have been found more effective than other methods based on second-order statistics or random fields, which analyze textures at a single resolution [5], [12]. Several approaches have been proposed to extract features in the wavelet domain with application in texture analysis. Often, the extracted texture features are wavelet energy signatures, which have been found useful for texture classification. The

second-order statistics of the wavelet transform (WT) have also been used to improve the accuracy of texture characterization [13]. Higher order dependencies of wavelet coefficients have been recently studied in the context of general statistical image modeling [14] and for texture analysis [5], [11], [15], [16].

A substantial amount of work has been done in texture analysis in the wavelet domain (e.g., focusing on textures found in nature, arts, or design). However, little attention has been given to industrial stochastic textures (e.g., textures of nonwoven textiles and paper). Therefore, most of the methods were proposed for texture analysis in general but were not designed for the specific problem of stochastic texture analysis in industry, where the texture visual appearance is similar to noise. For example, in some of the approaches mentioned above, feature extraction is carried out assuming subband independence at each resolution (e.g., [6], [10], and [12]), which is not verified experimentally, as discussed later. Other methods perform regular energy sampling in the wavelet domain (e.g., [17]), but regular patterns are rarely found in industrial stochastic textures. Also, the methods based on higher order statistics generally do not make explicit relevant stochastic texture features that are important for industrial applications (where process conditions are estimated based on specific texture parameters) (e.g., [5], [11], and [16]).

Some researchers have suggested explicit feature dimensions for texture discrimination [17], [18]. In fact, directionality (i.e., texture anisotropy) is useful for stochastic texture analysis, and some methods have been proposed for estimating local variability and anisotropy in stochastic textures at one particular resolution [2]. These methods are unable to capture texture stochastic features at various scales, which can be relevant for industrial stochastic texture discrimination.

In this paper, industrial stochastic textures are modeled as colored noise,[1] and a multiresolution scheme for industrial stochastic texture representation and analysis is proposed. We begin by describing how we model industrial stochastic textures and how we measure the image gradients at multiple resolutions using wavelets. Based on the wavelet decomposition, the grayscale distribution and texture anisotropy are measured at multiple resolutions. Next, a multiresolution distance measure for stochastic textures is introduced. Finally, we present some applications in industry, experimental results, and conclusions.

II. TEXTURE GRADIENTS IN MULTIPLE RESOLUTIONS USING WAVELETS

Our method relies on texture gradients and gray-level distributions at multiple resolutions. The main reason for emphasizing such features is that gradients and gray levels constitute supplementary measures for stochastic texture analysis at each resolution. It implies that we may assume the gradient features as independent of the gray levels and vice versa, which is a desirable feature in industry. For example, industrial materials such as nonwoven textiles and paper may be produced with different mass densities (thicknesses), i.e., different gray-

level statistics, but with similar spatial variability properties, i.e., similar formation anisotropy. Also, different industrial stochastic materials can present distinct mass density and spatial variability statistics.

To estimate the local gradients in multiple resolutions, we apply the undecimated two-dimensional WT proposed by Mallat and Zhong [21]. This WT requires calculation of two detail images, using one smoothing function $\phi(x, y)$ and two wavelets $\psi^i(x, y)$. The dilation of these functions are denoted by [21]

$$
\phi_s(x, y) = \frac{1}{s^2} \phi(\frac{x}{s}, \frac{y}{s})
$$
$$
\psi_s^i(x, y) = \frac{1}{s^2} \psi^i(\frac{x}{s}, \frac{y}{s}), \qquad i = 1, 2 \tag{1}
$$

and the dyadic WT $f(x, y)$, at a scale $s = 2^j$, has two detail components given by

$$
W_{2^j}^i f(x, y) = (f * \psi_{2^j}^i)(x, y), \qquad i = 1, 2 \tag{2}
$$

and one low-pass component given by

$$
S_{2^j} f(x, y) = (f * \phi_{2^j})(x, y). \tag{3}
$$

The coefficients $W_{2^j}^1 f(x, y)$ and $W_{2^j}^2 f(x, y)$ represent the details in the x and y directions, respectively. Thus, the image gradient at the resolution 2^j can be approximated by [14], [21]

$$
\mathbf{W}_{2^j} f(x, y) = \begin{pmatrix} W_{2^j}^1 f(x, y) \\ W_{2^j}^2 f(x, y) \end{pmatrix}. \tag{4}
$$

Since we are dealing with digital images $f[n, m]$, we use the discrete version of the WT [21], and the discrete wavelet coefficients are denoted in this paper by $W_{2^j}^i f[n, m]$, for $i = 1, 2$.

Local gradient magnitudes and orientations can be obtained at scale j based on $W_{2^j}^1 f[n, m]$ and $W_{2^j}^2 f[n, m]$, for $j = 1, 2, \ldots, J$ [14], [21]. In particular, the edge orientation will be used later in our discussion and is given by the gradient direction, which is expressed by

$$
\theta_{2^j} f[n, m] = \arctan \left(\frac{W_{2^j}^2 f[n, m]}{W_{2^j}^1 f[n, m]} \right). \tag{5}
$$

III. REPRESENTING INDUSTRIAL STOCHASTIC TEXTURES IN MULTIPLE RESOLUTIONS

Since it can be difficult to discriminate industrial stochastic textures visually, this paper focuses on stochastic texture features that could facilitate the texture interpretation in industrial applications. We are also interested in stochastic texture features that could be used to discriminate and classify distinct process conditions.[2] In other words, we are interested in the

---

[1]Estimation in colored noise fields using high-order statistics has been discussed by Mendel [19] and by Cohen and Francos [20].

[2]One or more samples can be used as reference images to indicate a particular process condition.
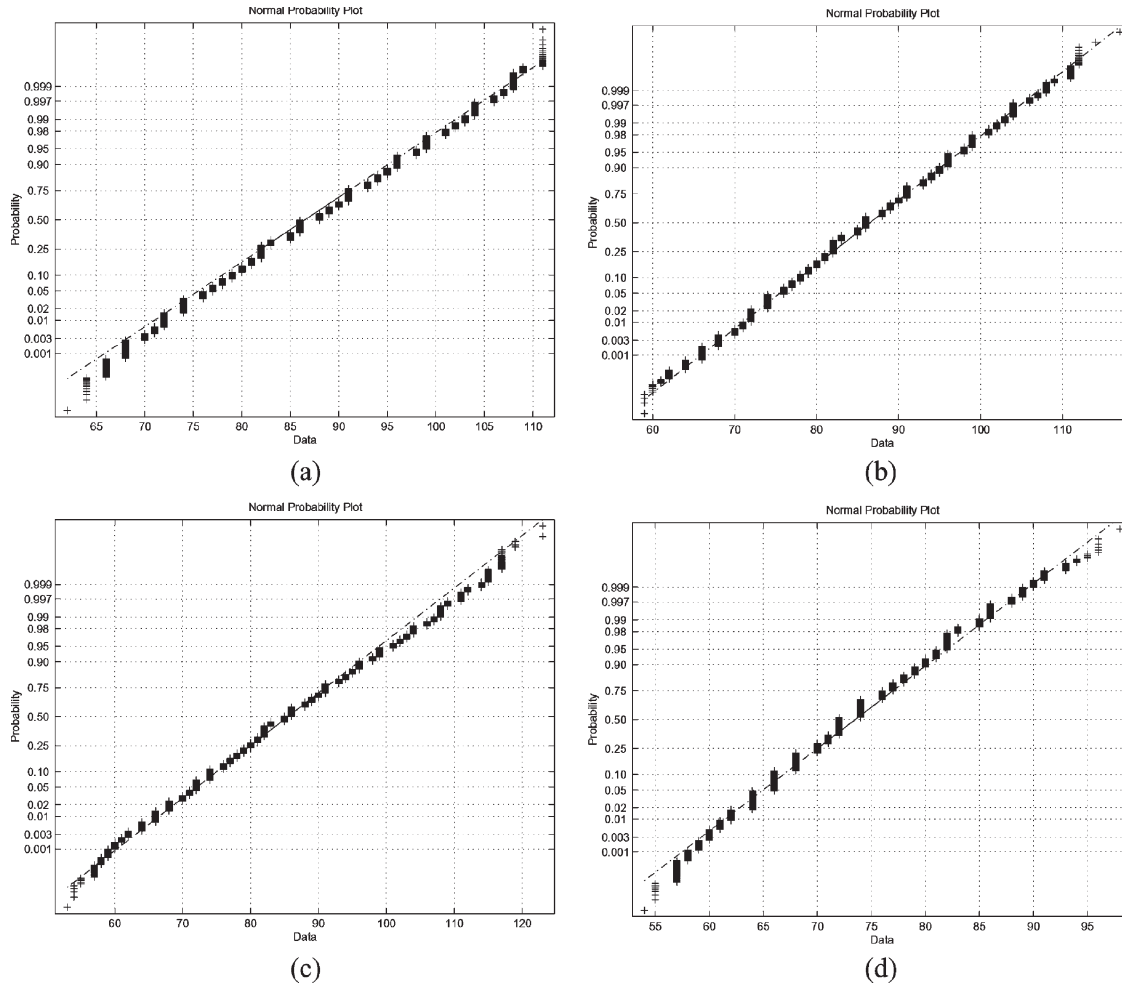
Fig. 1.   Normal plots of stochastic texture image examples (β-radiographs). (a) Headbox handsheet. (b) Sample from a Fourdrinier former. (c) Sample from a gap former. (d) Tissue paper sample.

scenario where the user provides an industrial stochastic texture image as a signature of a particular process condition, and a classification method returns the most similar image from a collection of images previously stored (i.e., images representing different process conditions). The stored and the user-provided texture images are represented by our texture features and are compared using a multiple resolution distance measure to be described later.

After applying the WT, we obtain the wavelet coefficients $S_{2^j}f[n,m]$, $W_{2^j}^1 f[n,m]$ and $W_{2^j}^2 f[n,m]$, for $j = 1, 2, \ldots, J$, where J is the number of scales. These wavelet coefficients are used to represent industrial stochastic textures at multiple resolutions, making explicit important texture features such as directionality and local gray-level variability.

Industrial stochastic textures such as nonwoven textiles and paper are textures characterized by nearly random gray-level spatial variability, which often present spatial correlations persisting over short, medium, or long ranges. The stochastic nature and correlation patterns of these textures originate in the stochastic fiber entanglements and alignments that take place during the nonwoven and paper-forming process [4].

The gray-level (i.e., local density) distribution in these textures can be modeled as a Gaussian process, with some de-

viation from normality at the end of the distribution tails [as illustrated in Fig. 1(a)–(d) using four industrial stochastic texture images].[3] In this paper, we model the stochastic texture patterns by Gaussian colored noise, where each texture pattern has a characteristic mean and spatial correlation [22].

We therefore model the input texture image $f(x, y)$ as Gaussian colored noise $\epsilon(x, y)$ with standard deviation σ and assume that this process is ergodic. Since the WT is a linear transform, the dyadic WT $\epsilon(x, y)$, or WT ϵ (using a simpler notation), at a scale $s = 2^j$, has two detail components given by

$$U_j^i = W_{2^j}^i \epsilon = \left(\epsilon * \psi_{2^j}^i \right), \qquad i = 1, 2 \qquad (6)$$

and one low-pass component given by

$$Z_j = S_{2^j} \epsilon = \left(\epsilon * \phi_{2^j} \right) \qquad (7)$$

---

[3]Only one set of four normal plots is illustrated in Fig. 1, but similar plots were observed in the other samples tested.

where $W_{2^j}^i \epsilon$ and $S_{2^j} \epsilon$ also are modeled as Gaussian colored noise processes, their standard deviations being $\sigma_{U_j^i}$ and σ_{Z_j}. The indexes $i = 1, 2$ correspond to the x direction and y direction, respectively.

Consequently, U_j^1 and U_j^2 are jointly Gaussian distributed with probability density function (pdf) [23], i.e.,

$$
p(U_j^1, U_j^2) = \frac{1}{2\pi \sigma_{U_j^1} \sigma_{U_j^2} \sqrt{1 - \rho_{12}^2}}
$$

$$
\cdot \exp\left[\frac{-1}{2(1 - \rho_{12}^2)} \left[\frac{(U_j^1)^2}{(\sigma_{U_j^1})^2} - 2\rho_{12} \frac{U_j^1 u_j^2}{\sigma_{U_j^1} \sigma_{U_j^2}} + \frac{(U_j^2)^2}{(\sigma_{U_j^2})^2} \right] \right]
$$

$$(8)$$

where ρ_{12} is the correlation coefficient of U_j^1 and U_j^2.

As before, since we are dealing with digital images $\epsilon[n, m]$, we use the discrete version of the WT [21], and the discrete wavelet coefficients are denoted in this paper by $W_{2^j}^i \epsilon[n, m]$, for $i = 1, 2$, and $S_{2^j} \epsilon[n, m]$.

A. Texture Directionality in Multiple Resolutions

Texture directionality, i.e., anisotropy, is an important parameter in the manufacture of foil-like materials. It correlates well with several mechanical and transport properties of such materials, as well as with commonly monitored manufacturing variables. It has already been shown that the distribution of angles $\theta_{2^j} \epsilon[n, m]$ over all pixels $[n, m]$ represents structural anisotropy [2]. In the case when all angles $\theta \epsilon$ are equally probable, the sample is isotropic.

In this paper, the occurrences of the coefficients $W_{2^j}^1 \epsilon[n, m]$ and $W_{2^j}^2 \epsilon[n, m]$ are approximated by Gaussian distributions. The normal plots in Fig. 2(c) and (d) show that the Gaussian model can represent a wide range of wavelet coefficient values in stochastic textures, with some deviation from Gaussian occurring in the tails of the distribution. This experimental evidence was confirmed in a large-sample test set. It was also verified that, in general, wavelet coefficients associated with the same image position $[m, n]$ in different subbands (i.e., $W_{2^j}^1 \epsilon[n, m]$ and $W_{2^j}^2 \epsilon[n, m]$) at the same resolution 2^j are correlated (i.e., are not independent). Therefore, we represent the joint distribution of wavelet coefficients by a bivariate Gaussian $G_{2^j}^{12}(W_{2^j}^1 \epsilon[n, m], W_{2^j}^2 \epsilon[n, m])$, denoted simply by G_j^{12}. The isoprobability curves of the bivariate Gaussian G_j^{12} are typically elliptic for anisotropic samples and tend to be circular for samples that are isotropic.

The joint coefficient distribution G_j^{12} determines two orthogonal axes of extremal variance that coincide with the directions of the eigenvectors $\mathbf{v}_{max}$ and $\mathbf{v}_{min}$ of the covariance matrix. To estimate the covariance matrix at resolution 2^j, denote the means of the coefficients $W_{2^j}^1 \epsilon$ and $W_{2^j}^2 \epsilon$ by μ_1 and μ_2, respectively. Due to the fact that signals have finite length, and also to the precision of the arithmetic, μ_1 and/or μ_2 is not necessarily zero. The covariance matrix is then calculated as follows:

$$
\mathbf{C}_{2^j} = E\left[\left(W_{2^j}^k \epsilon - \mu_k \right) \left(W_{2^j}^l \epsilon - \mu_l \right)^T \right]
$$

$$
= \begin{bmatrix} \sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 \\ \rho_{12} \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix} \tag{9}
$$

where $k = 1, 2$, $l = 1, 2$, ρ_{12} is the correlation coefficient of $W_{2^j}^1 \epsilon$ and $W_{2^j}^2 \epsilon$, and σ_1 and σ_2 are the standard deviations of $W_{2^j}^1 \epsilon$ and $W_{2^j}^2 \epsilon$, respectively.

To measure quantitatively the distribution eccentricity, we calculate the eigenvectors $\mathbf{v}_{max}$ and $\mathbf{v}_{min}$ from (9) and their corresponding eigenvalues λ_{max} and λ_{min}. The eigenvalues λ_{max} and λ_{min} define the semiaxes of a Gaussian ellipse aligned with the eigenvector directions. The eccentricity e of this ellipse is given by the ratio of the eigenvalues, i.e.,

$$
e = \frac{\lambda_{max}}{\lambda_{min}} \tag{10}
$$

and provides an estimate for the texture anisotropy at scale 2^j. For example, the measured eccentricity e is 1.008 for the isotropic sample in Fig. 2(a) and 1.227 for the anisotropic sample in Fig. 2(b).

It should be noticed that, at certain resolutions, the spatial texture image gradients may be higher than in others. This will be reflected in the covariance matrices with larger elements and will help in texture discrimination. The main orientation of the texture at resolution 2^j is given by the direction of $\mathbf{v}_{max}$, and the stronger the orientation (i.e., anisotropy) at 2^j, the more dominant is $\mathbf{v}_{max}$. Orientation may vary at different resolutions.

B. Texture Gray-Level Variability in Multiple Resolutions

The texture gray-level variability in multiple resolutions encodes important information about local density variability and, consequently, about the material structural homogeneity. In fact, local homogeneity is considered an important quality parameter of foil-like materials (e.g., nonwoven textiles and paper), namely the sample formation often referred in the specialized literature [4].

As discussed in Section III-A, the stochastic texture image coefficients $Z_j = S_{2^j} \epsilon$, for $j = 1, 2, \ldots, J$, are modeled as Gaussian colored noise [see the normal plots in Fig. 2(e) and (f)]. As a consequence, the distribution of coefficient values Z_j is represented by a Gaussian pdf, i.e.,

$$
G_j(Z_j) = \frac{1}{\sqrt{2\pi \sigma_{Z_j}^2}} e^{-0.5 \left[\frac{(Z_j - \mu_{Z_j})}{\sigma_{Z_j}} \right]^2} \tag{11}
$$

where μ_{Z_j} and σ_{Z_j} are the mean and standard deviation of the gray levels at scale 2^j, respectively.
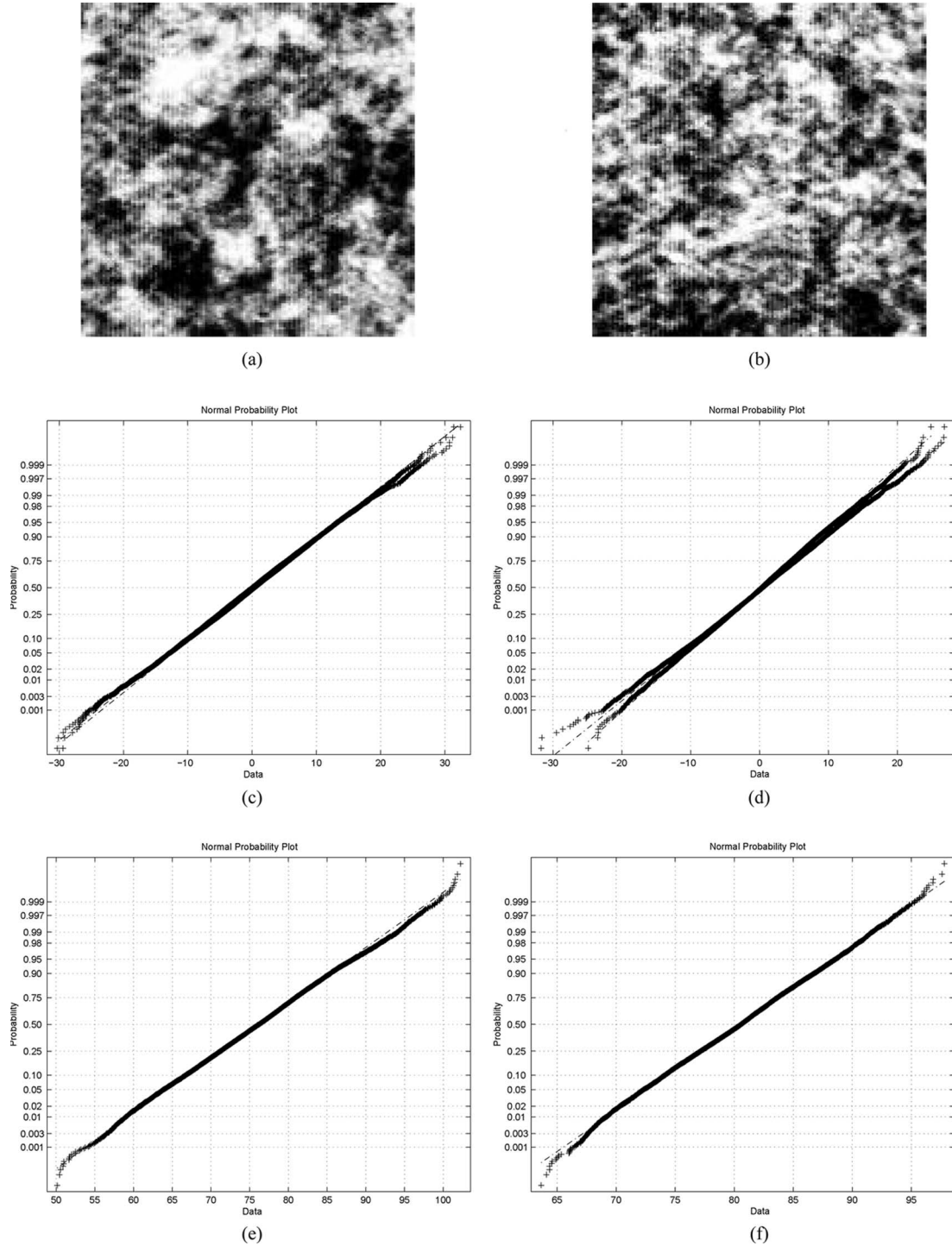
Fig. 2. Comparative results for isotropic and anisotropic test samples. Stochastic texture images (β-radiographs). (a) Nearly isotropic. (b) Anisotropic. Normal plots of coefficients W^i, $i = 1, 2$. (c) Nearly isotropic. (d) Anisotropic. Normal plots of coefficients S. (e) Nearly isotropic. (f) Anisotropic. Histogram equalization was applied to the images to improve their visualization.

Therefore, at each resolution, our stochastic texture representation consists of the bivariate Gaussian joint coefficient distribution G_j^{12}, encoding coefficient correlation and anisotropy information, and the univariate Gaussian distribution G_j, encoding gray-level variability, as illustrated in Fig. 3. Stochastic textures may vary in anisotropy, gray-level variability, or both, at different resolutions, and a wide range of industrial stochastic textures can be represented in terms of these features.

A distance measure to compare these texture representations is introduced next.

IV. CLASSIFICATION USING A DISTANCE MEASURE OF STOCHASTIC TEXTURES

The problem of classifying industrial stochastic texture images can be formulated as a multiple hypothesis (i.e., classes)
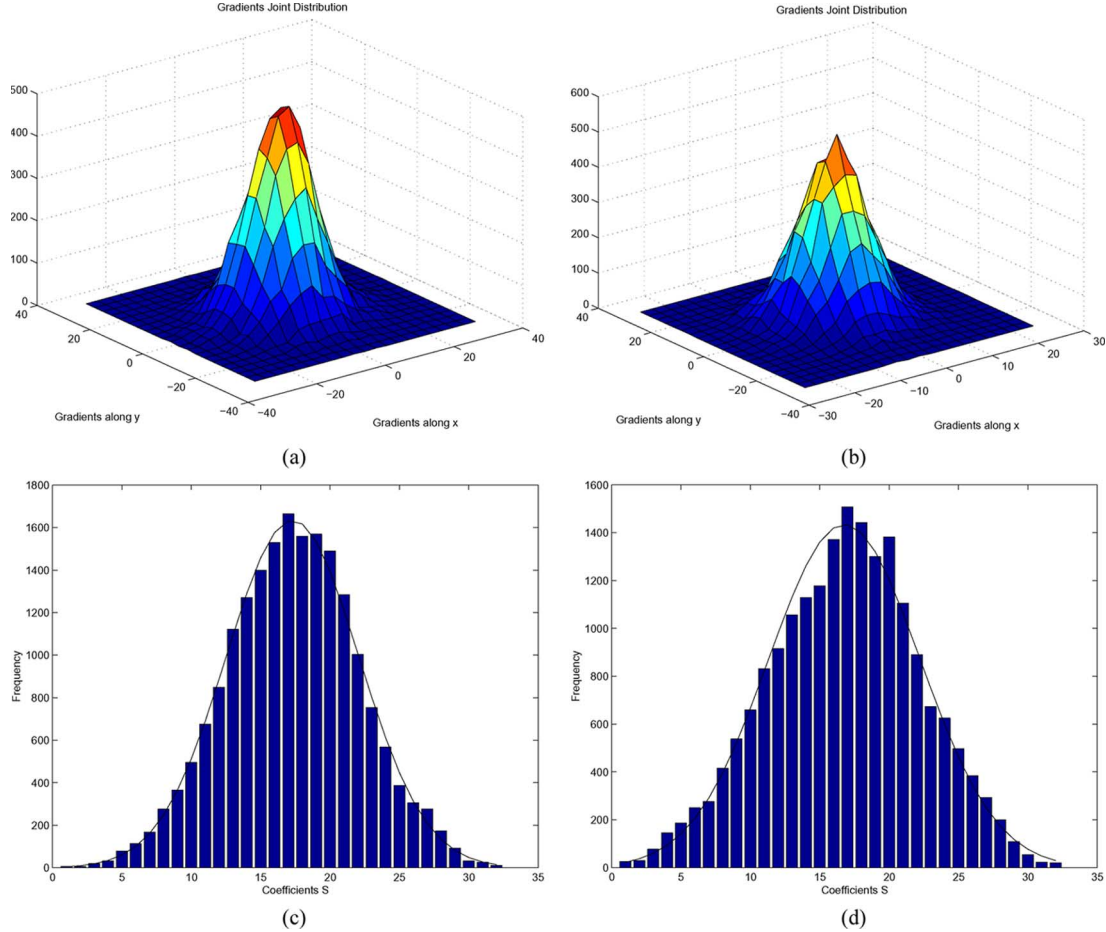
Fig. 3. Comparative results for isotropic and anisotropic test samples. Joint coefficient W^i, $i = 1, 2$ (gradient) value distributions. (a) Nearly isotropic. (b) Anisotropic. Coefficient S distributions. (c) Nearly isotropic. (d) Anisotropic.

testing problem, where we choose the best among T possible hypotheses $\{H^1, H^2, \ldots, H^t, \ldots, H^T\}$. In our approach, at each resolution 2^j, a stochastic texture image ϵ is represented in the WT domain by its coefficient distributions G_j^{12} and G_j. Under these circumstances, Do and Vetterli [10] showed that the Kullback–Leibler distance ranks the hypotheses H^t, or classifies the textures, consistently with the maximum-likelihood rule.

Our distance measure between stochastic textures therefore follows directly. The Kullback–Leibler distance between two bivariate Gaussian distributions $G(\mathbf{M}_1, \mathbf{C}_1)$ and $G(\mathbf{M}_2, \mathbf{C}_2)$ is given by [24]

$$D_{\mathrm{Gauss2}}\left(G(\mathbf{M}_1, \mathbf{C}_1), G(\mathbf{M}_2, \mathbf{C}_2)\right)$$
$$= \frac{1}{2}\log\left(\det(\mathbf{C}_2)/\det(\mathbf{C}_1)\right) + \frac{1}{2}\mathrm{trace}(\mathbf{C}_1\mathbf{C}_2^{-1})$$
$$+ (\mathbf{M}_1 - \mathbf{M}_2)^T\mathbf{C}_2^{-1}(\mathbf{M}_1 - \mathbf{M}_2) - \frac{n_d}{2} \quad (12)$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are the means of the bivariate distributions, $\mathbf{C}_1$ and $\mathbf{C}_2$ are the covariance matrices, and $\det(\mathbf{C})$ denotes the matrix determinant. Note that C_1 and C_2 are $n_d \times n_d$ matrices, and M_1 and M_2 are $1 \times n_d$ vectors (here, $n_d = 2$). On the other

hand, the Kullback–Leibler distance between two univariate Gaussian distributions $G(\mu_1, \sigma_1)$ and $G(\mu_2, \sigma_2)$ is

$$D_{\mathrm{Gauss}}\left(G(\mu_1, \sigma_1), G(\mu_2, \sigma_2)\right)$$
$$= \frac{1}{2}\left(\log\left(\frac{\sigma_2}{\sigma_1}\right)^2 + \left(\frac{\mu_1 - \mu_2}{\sigma_2}\right)^2 + \left(\frac{\sigma_1}{\sigma_2}\right)^2 + 1\right) \quad (13)$$

where μ_1, σ_1, μ_2, and σ_2 are the Gaussian parameters of the textures, respectively. In this paper, we use a symmetrical version of the Kullback–Leibler distance, i.e.,

$$D_{\mathrm{Gauss2}^s} = D_{\mathrm{Gauss2}}\left(G(\mathbf{M}_1, \mathbf{C}_1), G(\mathbf{M}_2, \mathbf{C}_2)\right)$$
$$+ D_{\mathrm{Gauss2}}\left(G(\mathbf{M}_2, \mathbf{C}_2), G(\mathbf{M}_1, \mathbf{C}_1)\right) \quad (14)$$
$$D_{\mathrm{Gauss}^s} = D_{\mathrm{Gauss}}\left(G(\mu_1, \sigma_1), G(\mu_2, \sigma_2)\right)$$
$$+ D_{\mathrm{Gauss}}\left(G(\mu_2, \sigma_2), G(\mu_1, \sigma_1)\right). \quad (15)$$

The Kullback–Leibler distance has some desirable properties. Its convexity guarantees that a minimum exists, and to calculate the Kullback–Leibler distance from multiple scales,

i.e., feature sets representing different wavelet subbands, we can use the "chain rule" [25], i.e.,

$$
\begin{aligned}
p(V_1, V_2, \ldots, V_n) &= p(V_1 | V_2 \ldots V_n) \\
&\times p(V_2 | V_3 \ldots V_n) \ldots p(V_{n-1} | V_n) p(V_n) \quad (16)
\end{aligned}
$$

where $p(V_1, V_2, \ldots, V_n)$ is a joint pdf with n marginal pdfs $p(V_1), \ldots, p(V_n)$.

It can be shown that the symmetrical Kullback–Leibler distance between two joint pdfs p^a and p^b

$$
\begin{aligned}
&D^s\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right) \\
&= D\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right) \\
&\quad + D\left(p^b(V_1, V_2, \ldots, V_n), p^a(V_1, V_2, \ldots, V_n)\right) \quad (17)
\end{aligned}
$$

has the average distance between marginals as a lower bound (see the Appendix for details), i.e.,

$$
\begin{aligned}
&D^s\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right) \\
&\geq \frac{1}{n} \sum_{i=1,\ldots,n} \left[D\left(p^a(V_i), p^b(V_i)\right) + D\left(p^b(V_i), p^a(V_i)\right) \right]. \quad (18)
\end{aligned}
$$

This distance lower bound provides a feature space contractive mapping and can be interpreted as a simpler version of the "full distance" (17). Samples not assigned to a given class c [i.e., because class c is not the "nearest" class according to the distance lower bound in (18)] will not be assigned to class c even if the full distance is used. In other words, the distance lower bound provides a simpler distance measure for sample classification and guarantees no false dismissals [26]. Later, the classification accuracy obtained by the distance lower bound will be discussed.

The variables $V_1, V_2, \ldots, V_n$ represent different wavelet subbands. If the same n subbands are used to classify all samples, the factor $1/n$ is a multiplying constant affecting equally all sample-to-class distances and may be dropped without changing the classification results. In this case, the joint Kullback–Leibler distance is approximated simply by the sum of n subband feature set distances, i.e.,

$$
\begin{aligned}
&D^s\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right) \\
&\simeq \sum_{i=1,\ldots,n} D\left(p^a(V_i), p^b(V_i)\right) + D\left(p^b(V_i), p^a(V_i)\right). \quad (19)
\end{aligned}
$$

Also, it shall be observed that different texture features do not necessarily carry the same weight in stochastic texture discrimination. Considering that $D(p^a, p^b) \geq 0$, relevances can be assigned to the different features V_i by imposing weights α_i to the feature set distances $D(p^a(V_i), p^b(V_i))$ in the distance

sum (deforming the feature space).[4] For example, suppose that two distinct feature sets are used for texture discrimination. If one of them is more relevant to achieve better texture discrimination, then its weight α_1 should be higher in the distance sum (i.e., $\alpha_1 > \alpha_2$). In this case, (19) is modified and written as

$$
\begin{aligned}
&D_p^s\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right) \\
&\simeq \alpha_1 \sum_{i=1,\ldots,n_1} \left[D\left(p^a(V_i), p^b(V_i)\right) + D\left(p^b(V_i), p^a(V_i)\right) \right] \\
&\quad + \alpha_2 \sum_{i=n_1+1,\ldots,n} \left[D\left(p^a(V_i), p^b(V_i)\right) + D\left(p^b(V_i), p^a(V_i)\right) \right] \\
&\simeq \alpha_1 D^s\left(p_1^a, p_1^b\right) + \alpha_2 D^s\left(p_2^a, p_2^b\right). \quad (20)
\end{aligned}
$$

Let us introduce our model in the context of the discussion above. As mentioned in Section II, at each scale, the feature sets associated with gray levels (i.e., Z_j) and with gradients (i.e., U_j^i, $i = 1, 2$) are approximated by orthogonal (i.e., independent) components of our stochastic texture representation. This approximation is justified on the basis of experimental evidence. Azimifar *et al.* [27] concluded that the majority of wavelet coefficients corresponding to random fields have correlations very close to zero (i.e., are uncorrelated) across and within scales, but some orientation subband coefficients are strongly correlated across scales or within a particular scale but across orientation subbands. In particular, in our experiments with industrial stochastic textures, we observed that scaling and orientation subband coefficients present very low correlation across and within scales (i.e., Z_j and U_j^i, $i = 1, 2$, are approximately uncorrelated).[5] In this case, a lower bound can be obtained in terms of $p(Z_j)$ and $p(U_j^{12})$, which are univariate and bivariate Gaussian distributions, respectively (see the Appendix for details).

Finally, the proposed multiresolution stochastic texture distance measure is

$$
D_j \equiv \sum_{j=1}^{J} \left((1-\alpha) D_{\text{Gauss2}^s}^j + \alpha D_{\text{Gauss}^s}^j \right) \quad (21)
$$

where $D_{\text{Gauss2}^s}^j$ and $D_{\text{Gauss}^s}^j$ are the Kullback–Leibler distances between bivariate Gaussian and univariate Gaussian distributions at each resolution j, and α ($\in [0, 1]$) is a parameter that controls the weights attributed to $D_{\text{Gauss2}^s}^j$ and $D_{\text{Gauss}^s}^j$. The weights of the gray-level and gradient representation components can be adjusted by the parameter α, and this parameter is chosen to maximize correct texture class discrimination. Therefore, at each resolution j, anisotropy is reflected in the distance between bivariate Gaussians $D_{\text{Gauss2}^s}^j$ (i.e., gradient orientations) and grayscale variability in the distance between univariate Gaussians $D_{\text{Gauss}^s}^j$ (i.e., gray-level distributions).

---

[4]In (19), all n individual features carry the same weight in the distance sum, and $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 1$.

[5]In our experiments, the maximum positive/negative correlation measured between Z and U^i, $i = 1, 2$, across scales was 0.001, and within the same scale, it was 0.004.

In the next section, we discuss some applications of our multiresolution stochastic texture representation, as well as the efficacy of our stochastic texture distance measure.

V. APPLICATIONS AND EXPERIMENTAL RESULTS

Industrial stochastic texture images are widely used in the manufacture of foil-like stochastic materials such as nonwoven textiles, paper, polymer membranes, conductor, and semiconductor coatings. In the case of nonwovens and paper, these materials are formed on a moving web by a process similar to the filtration of a fiber suspension, with some stochastic dynamics. Nowadays, to obtain higher productivity, the forming web tends to be several meters wide and moves along the machine direction at speeds of 400 m/min (or higher) [28]. Consequently, paper and nonwoven textiles have a definite "grain" caused by the greater orientation of fibers in the machine direction and by the stress/strain imposed during pressing and drying, in a later stage of manufacture. Therefore, the physical properties of such materials are usually anisotropic, and variations exist in the makeup of fibers and in fiber orientation along and across the moving web (i.e., along the moving web orthogonal direction, CD, and along the machine direction, MD) [28]. The ratio between physical properties measured along the MD and CD directions is often used in industrial laboratory quality tests and is known as the "MD-to-CD ratio."

Experienced machine operators often make an effort to use visual clues for grading samples, which are usually collected at different machine locations. For example, the spatial mass distribution in a sample (i.e., formation) is graded based on the sample texture visual appearance, estimating how "cloudy," "grainy," or "oriented" the sample texture looks. This is illustrated in Fig. 4(a) and (b), where sample 1 looks more cloudy than sample 29, sample 30 looks more grainy than sample 32, and samples 31–35 look more oriented than the other samples. A substantial amount of experience and intuition (subjective appearance evaluation) is used in this visual evaluation. During the past two decades, objective and automated testing of nonwoven materials has become a new trend in industry. An example of this new trend is the popular "TSI/TSO" orientation test [28], which estimates fiber orientation by measuring the sound speed along different sample directions. All the orientation tests mentioned above are destructive or require contact with the sample. Nondestructive and noncontact testing has been a challenge for researchers [1], [2], [4].

To test whether our texture representation is sensitive enough to capture forming condition fluctuations, which occur commonly in regular production, we used 20 sets of 11 samples equally spaced along CD (i.e., 15 cm^2 each, spaced 50 cm from each other). Ten sample sets represent standard operating conditions, and ten of them represent deviations from standard operating conditions (i.e., were obtained before stopping production for maintenance). The central parts of these samples were imaged using a scanner with transparency unit, obtaining images with 400×400 pixels, with nearly 0.3×0.3 mm^2/pixel. Three consecutive dyadic scales were used (2^j, for $j = 1, 2, 3$) for texture analysis. Our experiments aimed at evaluating web uniformity and ranking samples by their anisotropy. We estimated
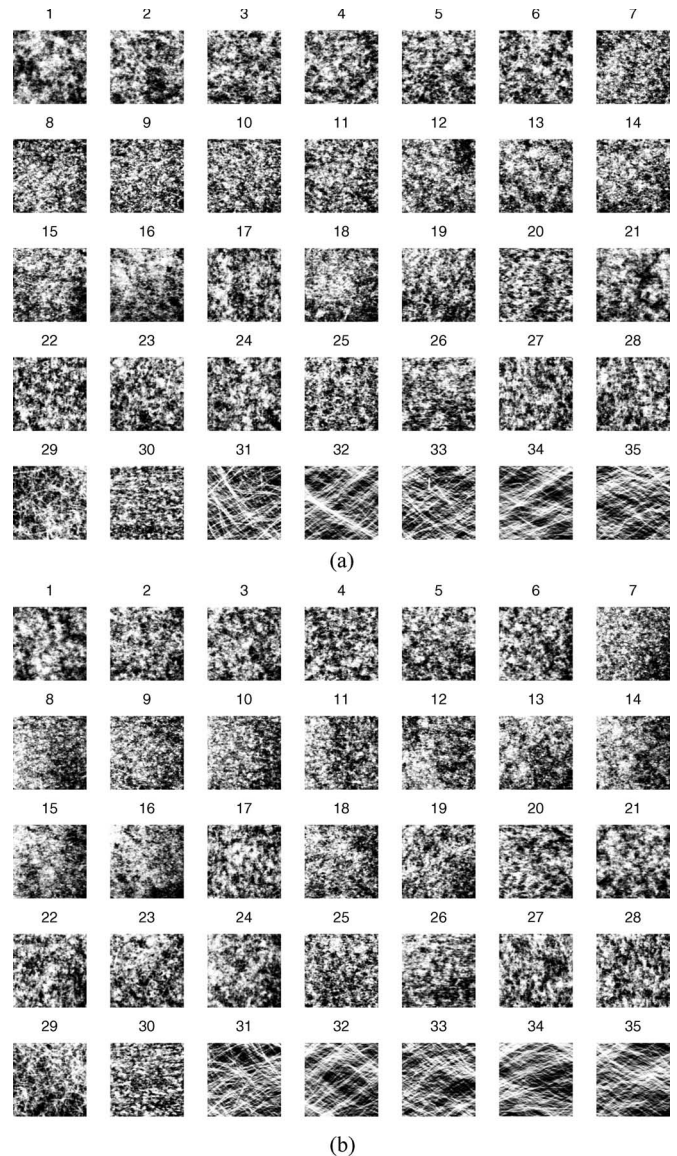

Fig. 4. Sample images of the 35 stochastic texture classes used in the experiments (β-radiographs). (a) Plate 1, showing one sample per class. (b) Plate 2, showing one sample per class (the samples are different of those displayed in Plate 1). Histogram equalization was applied to the images to improve their visualization.

anisotropy based on the middle scale (i.e., $j = 2$) and (10). We found that our approach tends to correlate well with anisotropy measurements based on the tensile test[6] (coefficient of correlation $r = 0.9059$). Also, we found that the sample rankings based on the anisotropy measured with our method and with the tensile test tend to correlate well ($r = 0.8844$). Our texture analysis correlates less well with TSI/TSO measurements along CD (i.e., $r = 0.7825$). For the stochastic texture classification experiments, we used a standard stochastic texture industrial database, as detailed next.

---

[6]The orientation of paper and nonwoven textiles can be determined by testing the mechanical resistance to tearing in the MD and CD (i.e., tensile strength test); the resistance to tearing usually is less in the MD.

A. Stochastic Texture Classification for Process Condition Analysis

To illustrate the performance of our approach in stochastic industrial texture classification, we used 315 distinct β-radiographic images of nonwoven textile and paper samples, obtained from a standard industrial image database [29]. We chose samples from a variety of forming machines, with different furnish and grammage[7] (e.g., among the 315 samples, we have samples of headbox handsheets, repulped machine sheets, standard handsheets, board handsheets, gap formers, Fourdrinier formers, speedformers, tissue papers, and glass-fiber mats). All these images have a resolution 140×140 pixels, with a spatial resolution of 0.2×0.2 mm^2/pixel [29]. We selected nine samples of each type, where each type of texture indicates a particular production condition (i.e., specific operating parameters and furnish), with a total of 35 classes (i.e., texture types). Typical samples of the 35 classes are displayed in Fig. 4. The sample textures representing different operating conditions were classified, and the system condition could be identified. Potentially, this approach could help nonwoven textile industry operators identify the current system condition, instead of relying on *ad hoc* stochastic texture interpretation methods.

In our classification experiments, we compared our approach, denoted here as $G2G$, KL [i.e., (21)], with the representation of wavelet coefficients by generalized Gaussians (i.e., GG, KL), using the optimal multiresolution texture distance proposed by Do and Vetterli [10].[8] The other approaches tested were as follows: 1) $G2G$, $Euclid.$, i.e., bivariate and univariate Gaussians used as in our approach, with the Euclidean distance applied to a seven-dimensional parameter space representing the Gaussians; 2) $GHist$, $Euclid.$, i.e., a Gaussian univariate modeling the gray levels, using the Euclidean distance applied to a two-dimensional parameter space representing the Gaussians; and 3) other approaches, such as Gabor filters and gray-tone spatial dependence (GTSD) matrices. In our problem, the classification accuracies obtained with Gabor filters and GTSD matrices were poorer than $G2G$, $Euclid.$, and their results are not considered further in our discussion. Table I shows the percentage of correct classification obtained experimentally, indicating that our approach gives the best correct classification rate for the 315 samples (and 35 classes) using $\alpha = 0.3$ (i.e., the best classification results were obtained with $\alpha = 0.3$). The other stochastic texture representation and classification approaches performed less well in tests, as shown in Table I. Therefore, besides providing an explicit representation of important stochastic texture features for texture analysis (e.g., anisotropy and local gray-level variability), our approach offers good performance compared with other approaches for texture representation and classification.

Our simplified distance measure of stochastic textures shown in (21) applies the distance lower bound concept, as discussed

TABLE I
MEASURED CORRECT CLASSIFICATION RATE FOR THE 315 SAMPLES BELONGING TO THE 35 TEXTURE CLASSES ILLUSTRATED IN FIG. 4, INTEGRATING THREE SCALES: $G2G$, KL IS OUR APPROACH, I.E., BIVARIATE AND UNIVARIATE GAUSSIANS AND KULLBACK–LEIBLER DISTANCE; GG, KL IS THE APPROACH OF GENERALIZED GAUSSIANS AND KULLBACK–LEIBLER DISTANCE [10]; $G2G$, $Euclid.$ IS THE APPROACH OF BIVARIATE AND UNIVARIATE GAUSSIANS AND EUCLIDEAN DISTANCE; $GHist$, $Euclid.$ IS THE APPROACH OF GAUSSIAN MODELS FOR GRAY LEVELS AND EUCLIDEAN DISTANCE

Classif. Approach	Correct Classif. Rate
G2G, KL	0.94
GG, KL	0.91
G2G, Euclid.	0.76
GHist, Euclid.	0.68

earlier. Recall that the full Kullback–Leibler distance D_{full}^s (17) considers the correlations between subband features at scales j_1 $(\in [1, J])$ and j_2 $(\in [1, J])$, i.e., $\rho(Z_{j_1}, Z_{j_2})$, $\rho(Z_{j_1}, U_{j_2}^1)$, $\rho(Z_{j_1}, U_{j_2}^2)$, $\rho(U_{j_1}^1, U_{j_2}^1)$, $\rho(U_{j_1}^2, U_{j_2}^2)$, and $\rho(U_{j_1}^1, U_{j_2}^2)$. However, our distance lower bound D_{LB}^s (21) is obtained by simplifications of D_{full}^s: 1) the feature sets $U^{1,2}$ and Z are assumed to be uncorrelated because they generally have very small correlation; and 2) the feature sets Z, U^1, and U^2 are assumed to be uncorrelated across scales, i.e., their across-scale correlations are discarded because $\rho(Z_{j_1}, Z_{j_2})$, $\rho(U_{j_1}^1, U_{j_2}^1)$, and $\rho(U_{j_1}^2, U_{j_2}^2)$, $j_1 \neq j_2$, are strong and do not vary much for stochastic textures in general, providing little discrimination information. Therefore, $D_{\text{full}}^s > D_{\text{LB}}^s$ is expected. We conducted experiments using the set of 315 stochastic textures to clarify the relationship between the full distance D_{full}^s and the distance lower bound D_{LB}^s. These experiments confirmed that the relationship between D_{full}^s and D_{LB}^s varies for different stochastic texture samples (i.e., changing U^1, U^2, Z, and their correlations), as expected from our theoretical developments. We measured strong across-scale correlations for the same subband features $\rho(Z_{j_1}, Z_{j_2})$, $\rho(U_{j_1}^1, U_{j_2}^1)$, and $\rho(U_{j_1}^2, U_{j_2}^2)$ (i.e., 0.94 for Z, and 0.81 for U^1 and U^2, in average; the measured standard deviations of the correlation values were 0.015 and 0.03, respectively). However, at the same scale j, we measured small correlations $\rho(U_j^1, U_j^2)$ (i.e., 0.01 in average) and even smaller correlations $\rho(Z_j, U_j^1)$ and $\rho(Z_j, U_j^2)$ (i.e., 0.002 in average).[9] In all our experiments, $D_{\text{full}}^s > D_{\text{LB}}^s$, and the ratio $D_{\text{full}}^s / D_{\text{LB}}^s$ varied from a minimum of 48.55 to a maximum of 2935.5 for different textures, reaching 289.79 in average. Our approach obtains correct classification in 94% of the 315 samples tested (belonging to the 35 classes), as indicated in Table I, and produces no false dismissals. However, the full distance may discriminate more accurately the 6% of the texture samples that are incorrectly classified by our simplified approach, at a higher computational cost.

B. Evaluation of the Forming Condition Homogeneity Across the Web

Given a set of industrial stochastic texture images, the distances between pairs of textures using (21) can be described by a gamma pdf, as detailed next. The family of gamma pdf is given by

$$\left\{ p(x; \tau, \nu) = \left(\frac{\nu}{\tau}\right)^{\nu} \frac{x^{\nu-1}}{\Gamma(\nu)} e^{-\frac{\nu}{\tau} x} | \nu, \tau \in R^{+} \right\}, \qquad x \in R^{+} \tag{22}$$

so the space of parameters is topologically $R^{+} \times R^{+}$. The gamma pdf is a member of the exponential family, and it includes as a special case ($\nu = 1$) the exponential distribution itself. Recently, Hwang and Hu [30] proved for $n \geq 3$ independent positive random variables x_i, with a common continuous pdf g and $i = 1, 2, \ldots, n$, that having the sample means $\bar{x}_i$ and the sample coefficients of variation $\text{cv}_i = S_{x_i}/\bar{x}_i$ independently distributed implies that g is a gamma distribution.[10] This is equivalent to saying that given n data sets collected independently, and represented by the variables $x_1, x_2, \ldots, x_n$, the underlying distribution g is gamma if the sample means $\bar{x}_i$ and sample standard deviations S_{x_i} are linearly related (i.e., present a high correlation coefficient). Now, considering n sets of stochastic texture samples obtained independently, and represented by their intraset sample distances $\{d_1\}, \{d_2\}, \ldots, \{d_n\}$, the common distribution of intraset sample distances g_d is gamma if the set distance means $\bar{d}_i$ are linearly related to the set distance standard deviations S_{d_i}. The adequacy of the g_d gamma model can be improved when the number of scales J is increased. The 35 classes of stochastic textures used in our experiments are represented in Fig. 5(b), which shows that intraclass sample distance means and standard deviations can be considered as linearly related (i.e., $r = 0.99$).[11] Therefore, for each class p, we assume that intraclass sample distance distribution is gamma and represent this distribution by the gamma parameters ν_p and τ_p.

In day-to-day industrial practice, production and systems maintenance are performed constantly. However, relatively low cost and objective methods of stochastic texture analysis for evaluating the forming conditions across the forming web are now in demand. The early identification of changes in production machine parameters and the early detection of faulty operating conditions provide important information for diagnosing system faults and for maintenance. These manufacturing systems are stochastic, and their fluctuations and irregularities are constantly monitored by sampling and testing at several locations across the forming web [28]. Our approach for stochastic texture representation can be used to evaluate the forming condition uniformity in manufacturing machines, as detailed next.

Fig. 5. Classification and web uniformity testing results. (a) Plot of the parameters (ν, τ) (dots: homogeneous/normal operation; crosses: inhomogeneous/abnormal operation). (b) Distance mean $\times$ standard deviation (coefficient of correlation = 0.99).

Homogeneous forming conditions across the forming web generate homogeneous texture sample sets, whose intersample distance distributions are clustered (i.e., the samples collected across the web present small statistical differences from each other and form a cluster in feature space). This can be evaluated from the point of view of the gamma family, where the parameter ν indicates clustering if $\nu < 1$, and ν indicates randomness if $\nu \simeq 1$ (i.e., when the samples collected across the web present significant statistical differences and do not form a cluster in feature space).[12] The homogeneity of the set of collected texture samples can be evaluated based on ν and τ parameters of the intersample distance distribution. Similar conclusions apply to the uniformity of the forming conditions used to produce such samples. Fig. 5(a) illustrates this concept, showing a plot of the parameters (ν, τ) for 20 distinct operating conditions. Each forming condition is represented by a set of

---

[10]In this case, $\text{cv}_i = S_{x_i}/\bar{x}_i = c$, where c is a constant.

[11]Cross-tabulating $\text{cv}_i \times \bar{x}_i$, it was verified that $p(\text{cv}_i, \bar{x}_i) \simeq p(\text{cv}_i) \cdot p(\bar{x}_i)$. The chi-square test for independence of lines and columns resulted in $\chi^2 = 26.5$ and a test significance level $p_s = 0.33$, suggesting independence of lines and columns.
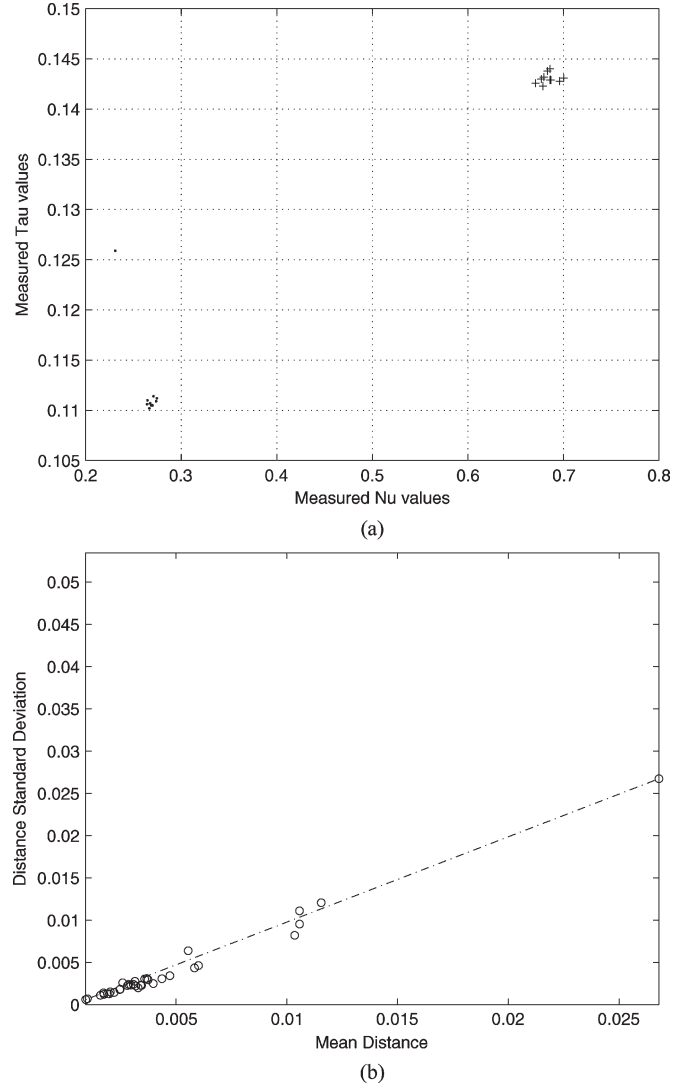
[12]The parameter $\nu > 1$ indicates dispersion in feature space, which is not relevant to our present discussion.

11 samples equally spaced, taken across the forming web; ten of these sets correspond to standard operating conditions, and ten of them indicate nonstandard operating conditions (adding up to a total of 220 samples). The nonstandard data have ν closer to 1 and are represented by "crosses" in Fig. 5(a). In this case, the sample inhomogeneity is substantial, which is indicated by the high variability of intersample distances (i.e., their low degree of clustering with $\nu \to 1$). The data interpretation suggests nonuniform operating conditions across the forming web, and that maintenance is presently required. On the other hand, the standard data sets are represented by "dots" in Fig. 5(a) and showed a high degree of intersample distance clustering, characterized by $\nu \ll 1$ (in our experiments, $\nu \leq 0.3$). Therefore, our experiments indicate the potential of using the parameters (ν, τ) as a tool to discriminate objectively between regular and deviant operating conditions during systems operation, at a relative low cost.

VI. Concluding Remarks

To summarize, stochastic texture images are acquired in large quantities in continuous industrial processes and encode important quality and process information. Moreover, the texture analysis methods required by some industrial processes must be objective because, often, texture discrimination exceeds the limits of human visual perception. Therefore, analyzing and classifying industrial stochastic textures are often challenging tasks.

The preliminary experimental results obtained by our approach were encouraging. Our stochastic texture representation can lead to new industrial applications, such as: 1) stochastic texture classification in industrial problems, where our method showed higher accuracy than other approaches, and 2) low-cost evaluation of production machine calibrations using stochastic texture images collected across the forming web. Our texture features tend to correlate well with industrial laboratory (manual) tests. However, more extensive testing is needed to achieve higher confidence levels in noncontact methods for obtaining anisotropy estimates, similar to those obtained in physical laboratory tests.

As future work, we intend to develop further our anisotropy detection approach to explore the dynamics of formation anisotropy. Also, we intend to study the influence of different scale weighting schemes and use our approach as a support for data mining operations in stochastic data repositories, with application in preventative maintenance and personnel training.

Appendix
Kullback–Leibler Distance Lower Bounds for Joint PDFs

The Kullback–Leibler distance between two joint pdfs, i.e., p^a and p^b, is given by [10]

$$D\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right)$$
$$= D\left(p^a(V_1|V_2 \ldots V_n), p^b(V_1|V_2 \ldots V_n)\right) + \cdots$$
$$+ D\left(p^a(V_{n-1}|V_n), p^b(V_{n-1}|V_n)\right) + D\left(p^a(V_n), p^b(V_n)\right). \tag{23}$$

Exchanging the variables $V_1, V_2, \ldots, V_n$ in n circular shifts, we obtain n joint probability functions that are equivalent in their representation of joint events: $p^\Psi(V_1, V_2, \ldots, V_n), p^\Psi(V_n, V_1, \ldots, V_{n-1}), \ldots, p^\Psi(V_n, V_{n-1}, \ldots, V_1)$, where $\Psi = a, b$. Now, adding the n Kullback–Leibler distances between joint probability functions p^a and p^b whose shifted variables correspond, and then rearranging terms, gives

$$D\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right)$$
$$= \frac{1}{n}\left[D\left(p^a(V_1|V_2 \ldots V_n), p^b(V_1|V_2 \ldots V_n)\right) + \cdots\right.$$
$$+ D\left(p^a(V_{n-1}|V_n), p^b(V_{n-1}|V_n)\right)$$
$$+ D\left(p^a(V_n), p^b(V_n)\right) + \cdots$$
$$+ D\left(p^a(V_n|V_1 V_2 \ldots V_{n-1}),\right.$$
$$\left. p^b(V_n|V_1 V_2 \ldots V_{n-1})\right) + \cdots$$
$$+ D\left(p^a(V_{n-2}|V_{n-1}), p^b(V_{n-2}|V_{n-1})\right)$$
$$+ D\left(p^a(V_{n-1}), p^b(V_{n-1})\right) + \cdots + \cdots$$
$$+ D\left(p^a(V_2|V_3 V_4 \ldots V_n V_1)\right.,$$
$$\left. p^b(V_2|V_3 V_4 \ldots V_n V_1)\right) + \cdots$$
$$+ D\left(p^a(V_n|V_1), p^b(V_n|V_1)\right)$$
$$\left. + D\left(p^a(V_1), p^b(V_1)\right)\right]. \tag{24}$$

Since $D(p^a, p^b) \geq 0$, we obtain

$$D\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right)$$
$$\geq \frac{1}{n}\left[D\left(p^a(V_1), p^b(V_1)\right) + D\left(p^a(V_2), p^b(V_2)\right) + \cdots\right.$$
$$\left. + D\left(p^a(V_n), p^b(V_n)\right)\right]. \tag{25}$$

Introducing the symmetric Kullback–Leibler distance in (23) and (25), these are written respectively as

$$D^s\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right)$$
$$= D\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right)$$
$$+ D\left(p^b(V_1, V_2, \ldots, V_n), p^a(V_1, V_2, \ldots, V_n)\right) \tag{26}$$
$$D^s\left(p^a(V_1, V_2, \ldots, V_n), p^b(V_1, V_2, \ldots, V_n)\right)$$
$$\geq \frac{1}{n}\left[D\left(p^a(V_1), p^b(V_1)\right) + D\left(p^b(V_1), p^a(V_1)\right)\right.$$
$$+ D\left(p^a(V_2), p^b(V_2)\right) + D\left(p^b(V_2), p^a(V_2)\right) + \cdots$$
$$\left. + D\left(p^a(V_n), p^b(V_n)\right) + D\left(p^b(V_n), p^a(V_n)\right)\right]. \tag{27}$$

In the context of this paper, the variables $V_{1,\ldots,n}$ represent U_j^1, U_j^2, and Z_j, where j is the WT scale of analysis. Rearranging the terms provided by the chain rule (16) and changing variables, we have

$$p\left(Z_1 U_1^1 U_1^2 \ldots Z_j U_j^1 U_j^2 \ldots Z_n U_n^1 U_n^2\right)$$
$$= p\left(Z_1 \ldots Z_j \ldots Z_n U_1^{12} \ldots U_j^{12} \ldots U_n^{12}\right) \tag{28}$$

where $U_j^{12} = U_j^1 U_j^2$. Applying the chain rule, we have

$$p\left(Z_1 U_1^1 U_1^2 \ldots Z_j U_j^1 U_j^2 \ldots Z_n U_n^1 U_n^2\right)$$
$$= p\left(Z_1 | Z_2 \ldots Z_j \ldots Z_n U_1^{12} \ldots U_j^{12} \ldots U_n^{12}\right)$$
$$\cdot \cdots$$
$$\cdot \; p\left(Z_n | U_1^{12} \ldots U_j^{12} \ldots U_n^{12}\right)$$
$$\cdot \; p\left(U_1^{12} | U_2^{12} \ldots U_j^{12} \ldots U_n^{12}\right) \ldots p\left(U_{n-1}^{12} | U_n^{12}\right) p\left(U_n^{12}\right).$$
$$(29)$$

Replacing the variables $Z \equiv Z_{i+1} \ldots Z_n$ and $U \equiv U_1^{12} \ldots U_n^{12}$ in (29) gives

$$p\left(Z_i | Z_{i+1} \ldots Z_n U_1^{12} \ldots U_n^{12}\right) = p(Z_i | ZU) = \frac{p(Z_i Z | U)}{p(Z | U)}$$
$$(30)$$

also, assuming that Z, Z_i are independent of U, we have

$$p(Z_i | ZU) \simeq \frac{p(Z_i Z)}{p(Z)} = p(Z_i | Z). \qquad (31)$$

Therefore, (29) is expressed as

$$p\left(Z_1 U_1^1 U_1^2 \ldots Z_j U_j^1 U_j^2 \ldots Z_n U_n^1 U_n^2\right)$$
$$= p(Z_1 | Z_2 \ldots Z_n) p(Z_2 | Z_3 \ldots Z_n)$$
$$\cdot \cdots$$
$$\cdot \; p(Z_{n-1} | Z_n) p(Z_n) p\left(U_1^{12} \ldots U_n^{12}\right)$$
$$= p(Z_1 \ldots Z_n) p\left(U_1^{12} \ldots U_n^{12}\right). \qquad (32)$$

Then, the symmetrical Kullback–Leibler distance is

$$D^s\left(p^a\left(Z_1 U_1^1 U_1^2 \ldots Z_n U_n^1 U_n^2\right), p^b\left(Z_1 U_1^1 U_1^2 \ldots Z_n U_n^1 U_n^2\right)\right)$$
$$= D^s\left(p^a(Z_1 \ldots Z_n), p^b(Z_1 \ldots Z_n)\right)$$
$$+ D^s\left(p^a\left(U_1^{12} \ldots U_n^{12}\right), p^b\left(U_1^{12} \ldots U_n^{12}\right)\right) \qquad (33)$$

and the lower bound is

$$D^s\left(p^a\left(Z_1 U_1^1 U_1^2 \ldots Z_n U_n^1 U_n^2\right),\right.$$
$$\left. p^b\left(Z_1 U_1^1 U_1^2 \ldots Z_n U_n^1 U_n^2\right)\right)$$
$$\geq \frac{1}{n} \sum_{j=1,\ldots,n} \left[D^s\left(p^a(Z_j), p^b(Z_j)\right)\right.$$
$$\left. + D^s\left(p^a\left(U_j^{12}\right), p^b\left(U_j^{12}\right)\right)\right] \qquad (34)$$

where $p(Z_j)$ and $p(U_j^{12})$ are univariate and bivariate Gaussian distributions, respectively. This lower bound tends to be an upper bound to (27), which only considers marginal probability distributions.

REFERENCES

[1] X. Z. Wang, *Data Mining and Knowledge Discovery for Process Monitoring and Control*. London, U.K.: Springer-Verlag, 1999.

[2] J. Scharcanski and C. T. J. Dodson, "Local spatial anisotropy and its variability," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 5, pp. 971–979, Oct. 2000.

[3] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Past, present and future," *J. Vis. Commun. Image Represent.*, vol. 10, no. 1, pp. 39–62, 1999.

[4] M. Deng and C. T. J. Dodson, *Paper: An Engineered Stochastic Structure*. Atlanta, GA: TAPPI, 1994.

[5] G. Fan and X. Xia, "Wavelet-based texture analysis and synthesis using hidden Markov models," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 50, no. 1, pp. 106–120, Jan. 2003.

[6] S. Arivazhagan and L. Ganesan, "Texture classification using wavelet transform," *Pattern Recognit. Lett.*, vol. 24, no. 9/10, pp. 1513–1521, Jun. 2003.

[7] S. W. Baik and P. W. Pachowicz, "Online model classification for adaptive texture recognition in image sequences," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 6, pp. 625–639, Nov. 2002.

[8] S. C. Zhu, Y. Wu, and D. Mumford, "Filters, random fields and maximum entropy," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 1–20, 1998.

[9] P. P. Raghu and B. Yegnanarayana, "Segmentation of Gabor filtered textures using deterministic relaxation," *IEEE Trans. Image Process.*, vol. 5, no. 12, pp. 1625–1636, Dec. 1996.

[10] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback–Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[11] H. Choi and R. G. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1309–1321, Sep. 2001.

[12] R. Manthalkar, P. K. Biswas, and B. N. Chatterji, "Rotation and scale invariant texture features using discrete wavelet packet transform," *Pattern Recognit. Lett.*, vol. 24, no. 14, pp. 2455–2462, Oct. 2003.

[13] G. V. Wouver, P. Sheunders, and D. V. Dyck, "Statistical texture characterization from wavelet representations," *IEEE Trans. Image Process.*, vol. 8, no. 4, pp. 592–598, Apr. 1999.

[14] J. Scharcanski, C. R. Jung, and R. T. Clarke, "Adaptive image denoising using scale-space consistency," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1092–1101, Sep. 2002.

[15] M. Clerc and S. G. Mallat, "The texture gradient equation for recovering shape from texture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 536–549, Apr. 2002.

[16] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1056–1068, Jul. 2001.

[17] N. Kim and S. Udpa, "Texture classification using rotated wavelet filters," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 30, no. 6, pp. 847–852, Nov. 2000.

[18] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 7, pp. 722–733, Jul. 1996.

[19] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.

[20] G. Cohen and J. M. Francos, "Least squares estimation of 2-D sinusoids in colored noise: Asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2243–2252, Aug. 2002.

[21] S. G. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 710–732, Jul. 1992.

[22] G. M. Jenkins and T. G. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.

[23] J. K. Patel and C. B. Read, *Handbook of the Normal Distribution*. New York: Marcel Dekker, 1982.

[24] S. Yoshizawa and K. Tanabe, "Dual differential geometry associated with the Kullback–Leibler information on the Gaussian distributions and its 2-parameter deformations," *SUT J. Math.*, vol. 35, no. 1, pp. 113–137, 1999.

[25] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1991.

[26] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases," in *Proc. ACM SIGMOD*, 1994, pp. 419–429.

[27] Z. Azimifar, P. Fieguth, and E. Jernigan, "Hierarchical Markov models for wavelet-domain statistics," in *Proc. IEEE Workshop Stat. Signal Process.*, 2003, pp. 258–261.

[28] G. A. Smook, *Handbook for Pulp and Paper Technologists*, 2nd ed. Vancouver, BC, Canada: Angus Wilde, 1994.

[29] C. Dodson, W. K. Ng, and R. R. Singh, *Paper Stochastic Structure Analysis—Archive 2*. Toronto, ON, Canada: Univ. Toronto, 1995. CD-ROM.

[30] T.-Y. Hwang and C.-Y. Hu, "On a characterization of the gamma distribution: The independence of the sample mean and the sample coefficient of variation," *Ann. Inst. Stat. Math.*, vol. 51, no. 4, pp. 749–753, 1999.
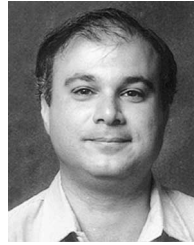
Jacob Scharcanski (M'03–SM'03) received the B.Eng. degree in electrical engineering and the M.Sc. degree in computer science from the Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil, in 1981 and 1984, respectively, and the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1993.

He was a Postdoctoral Fellow with the Pulp and Paper Centre, University of Toronto, Toronto, ON, where he worked on simulation of paper forming and on the analysis and interpretation of paper formation. He also was a Postdoctoral Fellow with the Communications Group, Department of Electrical and Computer Engineering, University of Toronto, where he worked on image processing and analysis. He is currently an Associate Professor with the Instituto de Informática, Universidade Federal do Rio Grande do Sul. He has authored or coauthored more than 70 publications in journals and conferences and has led to innovations in paper structural analysis by combination of image analysis and statistical geometry. His main areas of interest are image processing and analysis, pattern recognition, and industrial automation.