Stochasticity: A Feature for Analyzing and Understanding Textures, with Applications to Classification and Content-Based Image Retrieval

Abdourrahmane Atto, Yannick Berthoumieu, Rémi Mégret

▶ **To cite this version:**

Abdourrahmane Atto, Yannick Berthoumieu, Rémi Mégret. Stochasticity: A Feature for Analyzing and Understanding Textures, with Applications to Classification and Content-Based Image Retrieval. 2010. hal-00550318v1

HAL Id: hal-00550318

https://hal.science/hal-00550318v1

Preprint submitted on 26 Dec 2010 (v1), last revised 4 Nov 2013 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stochasticity: A Feature for Analyzing and Understanding Textures, with Applications to Classification and Content-Based Image Retrieval

Abdourrahmane M. ATTO[1], Yannick BERTHOUMIEU[2], Rémi MÉGRET[3]

[1,2] Université de Bordeaux, IPB, ENSEIRB-MATMECA,
IMS, CNRS UMR 5218, Signal and Image Group,
351 cours de la libération, 33405 Talence Cedex, France

***Abstract*—Stochasticity is proposed as a feature for texture characterization and analysis. Measuring stochasticity requires finding suitable representations that can significantly reduce statistical dependencies of any order. Wavelet packet representations provide such a framework for a large class of stochastic processes. The paper first addresses the selection of the best wavelet packet basis with respect to the stochasticity criterion and by using the Kolmogorov stochasticity parameter. A best basis under stochasticity consideration makes possible accurate texture description trough a dictionary of parametric models, especially for non regular textures. Among the properties of such a representation, the paper shows that texture classification is possible through stochasticity consideration. The relevance of the analysis also makes possible content-based stochasticity retrieval with semantics and with respect to the order structure of the wavelet packet bases.**

keywords: Texture Descriptors; Stochasticity Measurements; Semantic gap; Parametric modeling.

I. INTRODUCTION

The diversity of real world images has led researchers to use various mathematical tools in order to extract relevant image features or retrieve suitable information in images. For instance, probabilistic models, geometry properties and functional analysis have raised much dissertation in the last decades. In practice, the selection of appropriate features is driven by the class of images of interest. In this paper, the class of images we deal with can be conceptually defined through its departure from the class of regular images.

From the literature, a *regular* image is defined as either *smooth* or *geometrically regular* [1]: the image is composed of different smooth regions delimited by singularity curves. In contrast with the definition above, a *non-regular* image is such that: when splitting the image into smaller and smaller subimages (sub-surfaces), almost every subimage is non-regular in that it is expected to contain many delimitation curves.

[1] *abdou.atto@ims-bordeaux.fr*
[2] *yannick.berthoumieu@ims-bordeaux.fr*
[3] *remi.megret@ims-bordeaux.fr*

From the above consideration, a texture can be identified as either geometrically regular (composed of different or repetitive regions that are smooth except along their delimitation curves) or non-regular. Geometrically regular textures can be well characterized by local or global regularity measurements such as Holder exponents [1], [2], [3], [4], [5], [6], [7] or spectral measurements [8], [9], [10] [11]. In contrast, regularity measurements fail to be efficient for the characterization of non-regular textures since measuring very low regularity parameters is not straightforward.

The approach proposed below to characterize non-regular textures involves most relevant features that have proven useful in texture analysis [2]. These features are considered jointly in the framework of 1) *stochasticity*, a concept which relies on, but not limited to: coarseness, roughness and 2) wavelet packet transform, a representation that provides time-frequency, directionality, as well as other suitable statistical properties mentioned below.

The stochasticity degree (or randomness appearance) of a texture can be assessed by using the Kolmogorov stochasticity parameter [12]. This parameter applies under assumption that data are *independent and identically distributed* (*iid*) and their cumulative distribution function is completely specified. Recent works related to this parameter concern measuring the randomness degree of discrete sequences from dynamical systems and number theory [13], as well as measuring the contribution of randomness in the cosmic microwave background [14]. In these works, the Kolmogorov parameter has been used for specific datasets that are expected to comply with the underlying *iid* assumption, with a known distribution function.

In a more general framework involving real world textures, this *iid* assumption is very restrictive due to non-stationarity, spatial correlation and other more intricate statistical dependencies that occurs among texture samples. Moreover, the large variety of real world textures makes the use of a single family of distribution models inappropriate.

The contributions addressed in the present paper with respect to [13] and [14] concern being free of these restrictive assumptions by:

1) Considering the wavelet packet transform, a transform

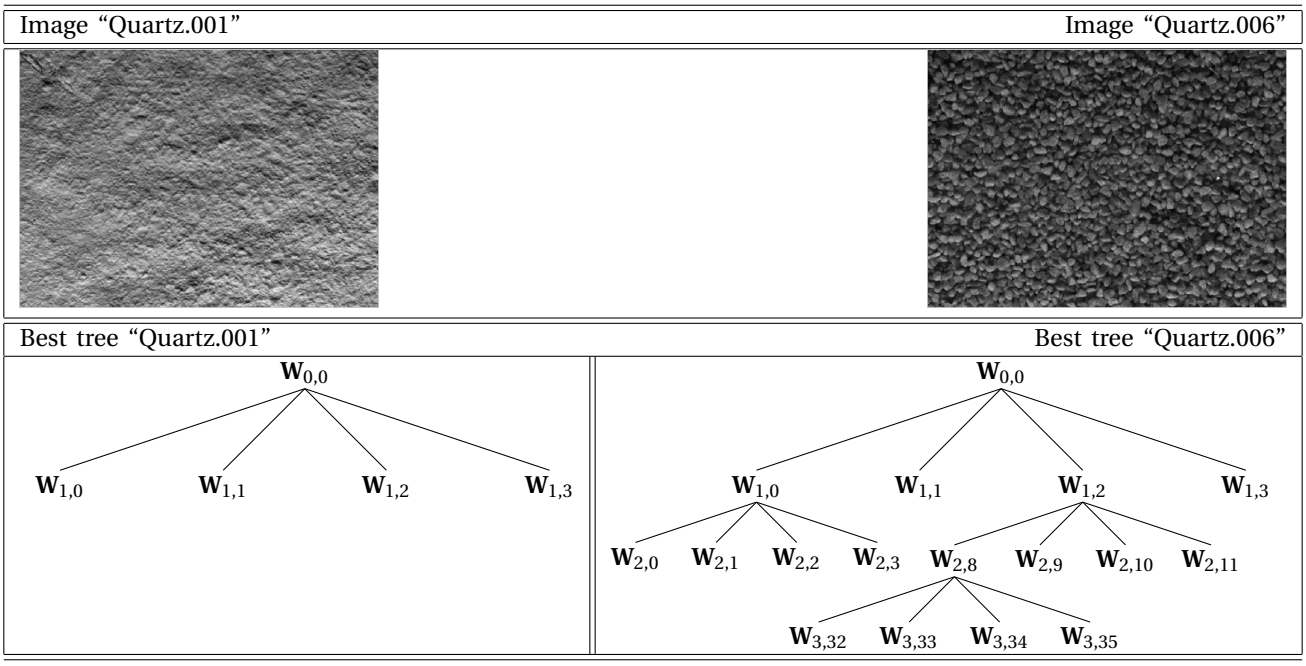

Fig. 1. Best representations of some OuTeX quartz textures depending on a fixed stochasticity bound: we split each node until the stochasticity parameters used have reached this bound. Stochasticity consideration makes it possible to characterize a texture by a specific tree structure in the wavelet packet domain.

that makes it possible to distribute many random processes as stationary, independent and identically distributed sequences, see for instance [15], [16], [17], [18];

2) Considering a dictionary of parametric models that are relevant with respect to the statistical distribution of the wavelet packet coefficients.

Under the above considerations, stochasticity can be measured even for correlated and non-stationary data through their wavelet packet representations. Figure 1 provides illustrative examples of texture descriptions with respect to their best stochastic wavelet packet bases: a texture is characterized by a tree structure whose terminal nodes are stochastic, in a sense to be highlighted further. Furthermore, from the order structure that characterizes wavelet packet bases, we derive texture classification by using the stochasticity degree and content-based texture retrieval with stochasticity semantics.

The presentation of this paper is as follows. Section II introduces bounds for stochasticity acceptance in natural images by using the uniform norm involved in the Kolmogorov stochasticity parameter. These bounds define a semantic template composed of 4 classes: *non-stochastic*, *quasi-stochastic*, *stochastic* and *strongly-stochastic*. Then, Section III provides insightful results on the algebraic structure on the set of wavelet packet bases. Section IV discusses stochasticity measurements in wavelet packet bases through model selection by using a dictionary of parametric probability distribution functions. Section V presents application of stochasticity analysis to the classification and content-based image retrieval for textured images. Finally, Section VI provides a conclusion to the paper.

II. Stochasticity Measurements

Let $X = (X_1, X_2, \ldots, X_N)$ be a sequence of absolutely continuous real random variables with respect to the Lebesgue measure on $\mathbb{R}$. This sequence is assumed to be *iid* with *probability density function, pdf* f and *cumulative distribution function, cdf* F.

The stochastic nature of a sample set $x = (x_1, x_2, \ldots, x_N)$ under the distribution model F is linked on how well we can consider this set as a realization of the discrete sequence X. This stochasticity can be assessed by using the Kolmogorov parameter [12]:

$$\lambda_N = \lambda_N(x, F) = \sqrt{N} \sup \left| F_{x,N}(t) - F(t) \right|, \tag{1}$$

where $F_{x,N}$ is the empirical *cdf* of the N-sample sequence x.

Testing stochasticity for random generators involves finding critical values (lower and upper bounds) for λ_N when N tends to ∞. These critical values can be computed from the Kolmogorov distribution: the upper bound on λ_N can be fixed by seeking a value C_0 such that $\Phi(C_0) \cong 1$ for stochasticity acceptance of the samples issued from this generator, with Φ being the Kolmogorov distribution defined as

$$\Phi(\lambda) = 1 + 2 \sum_{k=1}^{+\infty} (-1)^k e^{-2k^2\lambda^2}. \tag{2}$$

The rational of this approach is due to an asymptotic property of the sequence $(\lambda_N)_N$: when N tends to ∞, the *cdf* of the random variable λ_N converges to the Kolmogorov distribution Φ. Thus, asymptotically, the observation of a stochasticity value $\lambda_N \leqslant C_0$ is a certain event. In this sense, a random generator yielding $\lambda_N >> 2.4$ as the sample size size grows will be irrelevant.

In this paper, we are not concerned by the above asymptotic consideration for fixing critical values: we are interested in comparing two datasets in terms of their randomness degree. This can be performed according to their stochasticity parameters, whatever the values of these parameters. The question which arises is then: how far is the Kolmogorov parameter relevant, compared to other measures available from the literature on the topic? Section II-A provides some answers through an analysis of the distributions (*cdf* versus *pdf*) involved in stochasticity parameters and the norm used (cumulative versus uniform norms). Before presenting these results, we need to find an elementary non-stochastic pattern of reference. This pattern will be used to corrupt stochastic data and the sensitivity of different measures will then be addressed with respect to the behavior of these measures, when the size of the pattern increases.

From the Glivenko-Cantelli theorem, the quantity $\sup\left|F_{x,N}(t)-F(t)\right|$ decreases to 0 as $N\to\infty$. This implies that $\sup\left|F_{x,N}(t)-F(t)\right|$ is close to 0 for stochastic datasets. Note that $\sup\left|F_{x,N}(t)-F(t)\right|\leqslant 1$ so that any x satisfying $\sup\left|F_{x,N}(t)-F(t)\right|=1$ is necessarily non-stochastic. For instance, we will say that a constant sequence is deterministic with respect to any given continuous distribution model. Indeed, for such a sequence, we have that $\sup\left|F_{x,N}(t)-F(t)\right|=1$. Furthermore, we have that the presence of a value with large occurrence in a dataset can be qualified as a deterministic pattern since it impacts $\sup\left|F_{x,N}(t)-F(t)\right|$.

A. The relevance of the Kolmogorov stochasticity parameter in detecting deviations from a specified distribution

There are basically two criteria that distinguish stochasticity measures:

1.) the norm used, which can be cumulative or uniform.
2.) the distribution, which can be specified as a *pdf* or a *cdf*.

As examples, the Kolmogorov-Smirnov test [19] is based on the uniform (ℓ_∞) norm and compares two *cdfs*. The chi-square test [20] uses the cumulative ℓ_2 norm for comparing *pdfs*.

Let us consider a dataset having stochasticity degree η with respect to a given distribution model. Assume that these data are affected by a purely deterministic pattern in the sense that a proportion K/N of the data is set to a constant value. Since a stochasticity measure can be seen as a dissimilarity measure between distribution functions, then a relevant stochasticity measure is such that its stochasticity parameter should increases as K increases. In the following experiments, a purely deterministic pattern consisting in the insertion of K occurrences of a fixed value is introduced in datasets and we test the relevance of different stochasticity measures when the size K of this pattern increases.

Simulations are carried out both on data issued from random generators and on texture images. For convenience, we present the experiments concerning textured images. These experiments are performed upon the detail wavelet coefficients of the textures under consideration. These coefficients are expected to be very small in smooth regions and large in the neighborhood of edges. Increasing the number of null coefficients (if any) by forcing K large coefficients to zero (deterministic pattern) results in smoothing some edges of the image. This implies reducing the intrinsic stochasticity of the data when K increases. A relevant stochasticity measure should admit an increasing deviation from the initial stochasticity degree[1], when K increases.

For testing stochasticity measures, we consider the experimental setup presented in Table I. We use different combination between norms (ℓ_2,ℓ_∞) and distribution specification (*cdf*, *pdf*). In this table, $\|\cdot\|$ specifically denotes either the ℓ_∞ and ℓ_2 norms or the Kullback-Leibler[2] Divergence (KLD). In addition, if $c_{j,n}$ denotes the wavelet packet coefficients obtained at subband $\mathbf{W}_{j,n}$, then $c^K_{j,n}$ corresponds to the dataset obtained by setting the K largest values of $c_{j,n}$ to 0. In particular, $c^0_{j,n}=c_{j,n}$. The *Minimum Expected Value* (Min-EV) of the relative stochasticity value computed in Table I is the value obtained when no coefficients are set to zero: this value equals 1.

TABLE I
EXPERIMENTAL SETUP FOR TESTING THE RELEVANCE OF THE UNIFORM (ℓ_∞) NORM VERSUS THE CUMULATIVES ℓ_2 NORM AND KULLBACK-LEIBLER DIVERGENCE (KLD) IN STOCHASTICITY MEASUREMENTS. THE QUANTITIES INVOLVED IN THE COMPUTATION OF THE RELATIVE STOCHASTICITY VALUE (RSV) ARE THE EMPIRICAL DISTRIBUTION AND THE MODEL.

For $0\leqslant K\leqslant 150$, **do:**
- **Compute** the wavelet coefficients $\left(c_{j,n}\right)_{j,n}$ of the input image.
- **Introduce** a deterministic pattern among the coefficients of a subband by setting the K largest coefficients to zero (notation $\left(c^K_{j,n}\right)_{j,n}$).
- **Compute** the stochasticity parameters:
  Check the distribution type from variable "specification"
  - **Case** specification is "*cdf*", **then**:
    Compute the RSV: $\dfrac{\left\|F_{c^K_{j,n},N}-F_\theta\left(c^K_{j,n}\right)\right\|}{\left\|F_{c^0_{j,n},N}-F_\theta\left(c^0_{j,n}\right)\right\|}$,
  - **Case** specification is "*pdf*", **then**:
    Compute the RSV: $\dfrac{\left\|f_{c^K_{j,n},N}-f_\theta\left(c^K_{j,n}\right)\right\|}{\left\|f_{c^0_{j,n},N}-f_\theta\left(c^0_{j,n}\right)\right\|}$

End

Compare the measurements obtained: for a relevant measure, the larger K, the larger the stochasticity parameter.

The family of distribution functions chosen for modeling the wavelet coefficients is the *generalized gaussian* distributions. In addition, Gaussian, triangle and Epanechnikov

[1]The initial stochasticity degree is the value of the stochasticity parameter when no coefficients are forced to zero.

[2]The Kullback-Leibler similarity measure between random variables X_1 and X_2 having probability distribution functions f_{X_1} and f_{X_2} is defined as

$$\mathscr{K}(X_1,X_2)=\mathscr{K}(X_1\|X_2)+\mathscr{K}(X_2\|X_1),$$
$$\text{with}\qquad \mathscr{K}(X_i\|X_j)=\int_{\mathbb{R}} f_{X_i}(x)\log\frac{f_{X_i}(x)}{f_{X_j}(x)}\,dx,\quad i,j=1,2.$$

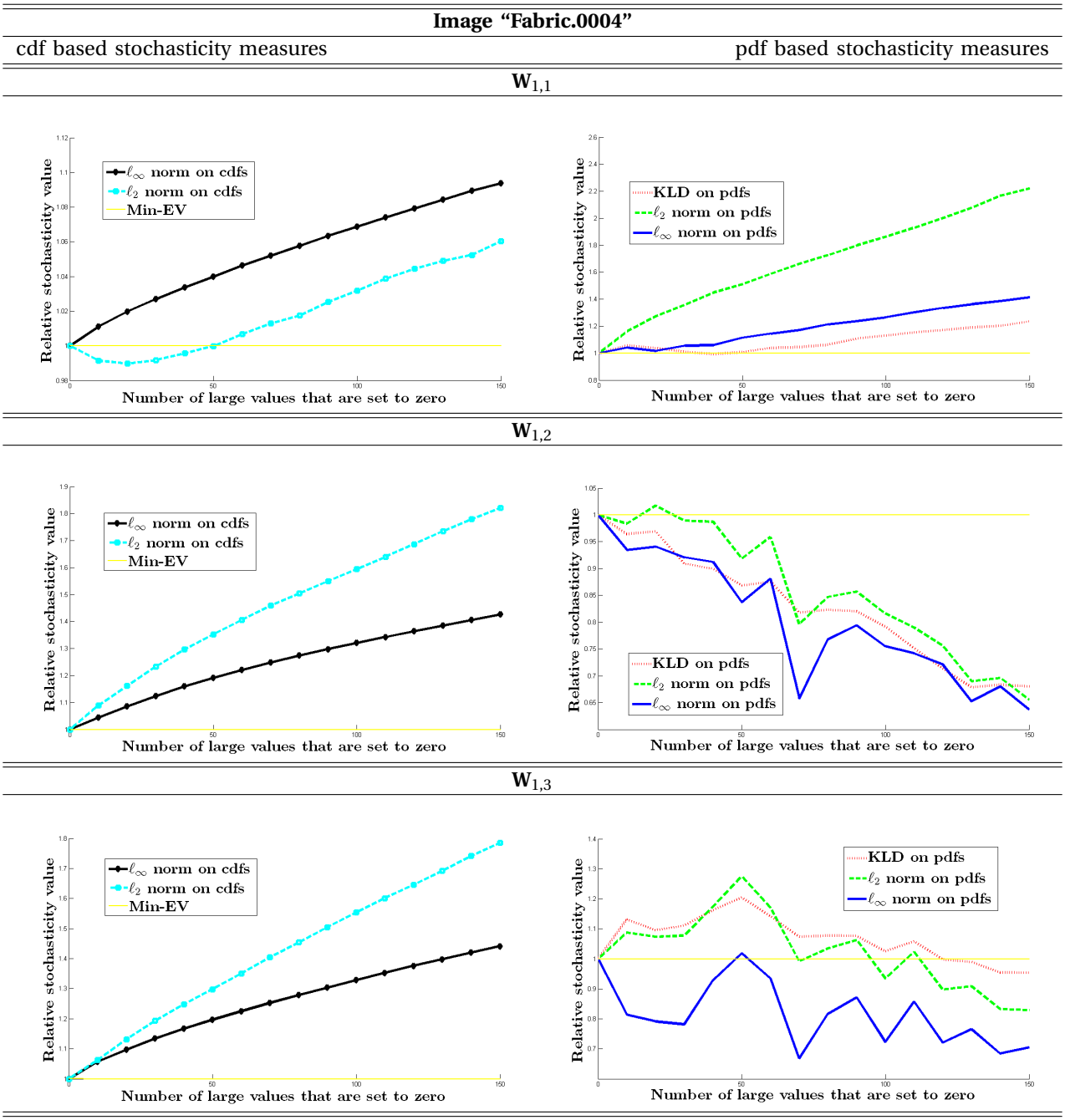

Fig. 2. Relative stochasticity values for the image "Fabric.0004" from the VisTeX database. A relevant stochasticity strategy should results in an increasing function starting from minimum expected value (Min-EV) 1 (represented by a solid line in the figures).

kernels have been used for *pdf* estimation. The results provided in Figures 2 and 3 are obtained with a Gaussian kernel and the wavelet decomposition has been performed with a Daubechies wavelet function of order 7. These results concern the images "Fabric.0004" and "Fabric.0018" from the VisTeX database. The results we have obtained are similar for other textures from the VisTeX database and for other kernels concerning *pdf* based measures.

As it can be seen in Figures 2 and 3, the uniform norm on the *cdfs* is the sole strategy that ensures an increase of the stochasticity parameter from its initial value, when the size K of the pattern increases. Cumulative measures (ℓ_2, KLD), as well as *pdf* based specifications are not very relevant for stochasticity assessment because of non-increasing deviations from the initial stochasticity degree: the local information is blurred through the averaging effect induced by cumulative measures or through the neighborhood consideration for computing *pdfs*.

Complementary tests performed on synthetic random numbers, without the use of wavelet transform, confirm the above conclusion. The results of these tests are not presented in order to ease the reading of the paper and also because of the limited size of the present paper. From now on, we assume that the stochasticity parameter

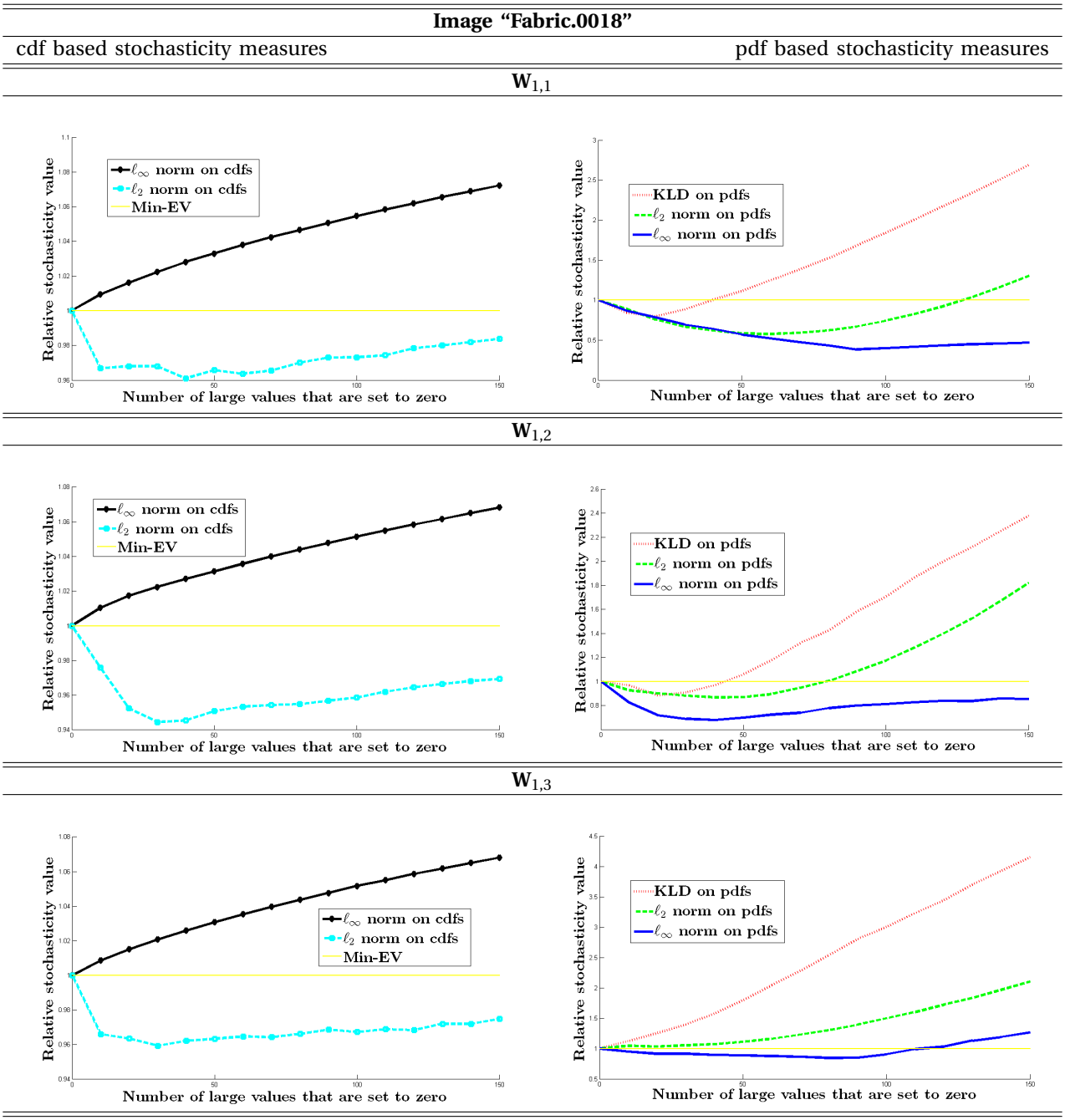

Fig. 3. Relative stochasticity values for the image "Fabric.0018" from the VisTeX database. A relevant stochasticity strategy should results in an increasing function starting from minimum expected value (Min-EV) 1 (represented by a solid line in the figures).

is of Kolmogorov's type: uniform norm that applies to compare the empirical *cdf* with the distribution model. Section II-B addresses the choice of different bounds on this parameter for generating a semantic stochasticity template. This makes it possible to classify textures relatively to the location of their stochasticity parameters on the stochasticity template.

B. Kolmogorov stochasticity measure versus *error-bounds from image estimation*

As shown in Section II-A, Kolmogorov parameter, given by Eq. (1), is consistent for detecting deterministic patterns in stochastic datasets and *vice versa*. For the main purpose of this paper, it is convenient to specify stochasticity bounds that make it possible to classify textures depending on their stochastic degrees. Note that the standard approach on stochasticity testing, [19], is binary and is based on asymptotic consideration.

In practical applications involving natural images, deriving a critical value on Eq. (1) from asymptotic consideration on the sample size is restrictive for assessing the intrinsic stochastic nature of the scene under consideration. Indeed, from the fixed sampling grid and quantification steps involved in image acquisition, natural images are with fixed

dynamical range. For instance, images are described by using integer values from 0 to 255 for a 8-bit coding, so that $\sup\left|F_{x,N}(t)-F(t)\right|$ is not expected to decrease significantly when the sample size increases[3].

In addition, note that this asymptotic consideration yields a binary decision making. This decision is taken by comparing $\sup\left|F_{x,N}(t)-F(t)\right|$ to $\mathsf{C}/\sqrt{N}$, where C is the critical upper bound derived from the Kolmogorov distribution Φ given by Eq. (2). It is worth emphasizing that such a decision is also restrictive for the "natural" perception of stochasticity in textured images: assume that, from a given texture (query) composed of N pixels, we build a larger image by copying periodically the query texture. We thus have the same texture, but two different stochasticity values for its characterization. These values are 1) the stochasticity value obtained from the query and 2) that obtained from the larger image. These values differ significantly since the sample term $\sqrt{N}$ has changed whereas $\sup\left|F_{x,N}(t)-F(t)\right|$ will remains approximately unchanged.

From the above analysis, the stochasticity degree of a given image is hereafter assessed from the height of the error measurement: $\sup\left|F_{x,N}(t)-F(t)\right|$. In contrast with the binary decision based on asymptotic consideration, we propose to derive different semantic classes consisting in stochasticity categories, by fixing bounds on $\sup\left|F_{x,N}(t)-F(t)\right|$ and with fixed sample size consideration. This is performed by dealing with $\sup\left|F_{x,N}(t)-F(t)\right| < \eta_i$ as a problem of estimating an unknown function from observed samples, and by fixing η_i so as to guarantee a PSNR greater than Ω_idBs, where $(\Omega_i)_i$ are bounds taken from standards on PSNR quality from image denoising and compression problems.

The PSNR (Peak Signal-to-Noise Ratio, in deciBel unit, dB) is given by

$$\mathrm{PSNR} = 10\log_{10}\left(d^2/\mathrm{MSE}\right), \tag{3}$$

where d is the dynamic of the image and MSE denotes the Mean Squared Error.

Proposition 1
Consider the problem of fitting $F_{x,N}(t_i)$ by $F(t_i)$ for $i = 1,2,\ldots,N$. Then in order to have a PSNR greater than ΩdBs, it suffices that $\eta^2 \leqslant d\times 10^{-\Omega/10}$.

The dynamic of $F_{x,N}$ is 1. Now, we set $\Omega_0 = 30$dBs, $\Omega_1 = 35$dBs and $\Omega_2 = 40$dBs (a minimum of 30dBs is required for an image denoising or compression algorithm to be relevant). These values are associated with the constants:

$$\eta_i = \sqrt{10^{-\Omega_i/10}}, \quad i = 0,1,2,$$

hereafter referred as lower bounds for high, good and fair quality set indicators. Then, we will use below, 4 stochasticity classes:

[3]for a random sequence x issued from distribution F, the decay of $\sup\left|F_{x,N}(t)-F(t)\right|$ as $N\to\infty$ follows from the Glivenko-Cantelli theorem.

Definition 1 (Semantic stochasticity classes)
Let

$$\kappa(x,F) = \sup_t \left|F_{x,N}(t)-F(t)\right|. \tag{4}$$

A sample set x is said to be strongly stochastic (resp. stochastic, quasi-stochastic, non-stochastic) with respect to a continuous *cdf* F if $\kappa(x,F)\in\left[0,\eta_2\right]$ (resp. $\kappa(x,F)\in\left]\eta_2,\eta_1\right]$, $\kappa(x,F)\in\left]\eta_1,\eta_0\right]$, $\kappa(x,F)\in\left]\eta_0,+\infty\right[$).

In order to measure stochasticity from a large class of stochastic processes including non-stationary and non-*iid* processes, we need the two properties:

(P1) a transform that has stationarization, decorrelation and higher order dependency reduction for the class of stochastic processes considered.

(P2) a large class of *cdfs* so as to approach with a good precision an arbitrary empirical *cdf* by an element of the class.

From [21], [22], [23], [24], [25], [18], [17], [16], [15], among others, wavelet packet bases have property **(P1)** when the decomposition level increases, provided that the filters used are with high orders. In order to have Property **(P2)**, we need to build a large dictionary composed of several families of *cdfs* so as to fit with any arbitrary continuous distribution shape. Fortunately, the statistical properties of the wavelet packet coefficients make it possible to drastically reduce the number of relevant distribution families involved in the dictionary. The next sections address the best wavelet packet basis selection with respect to stochasticity measurements. We begin by presenting, in the next section, some results on the algebraic structures that characterize the wavelet packet bases. These structures will be useful in the sequel.

III. Preliminary results on order structures among wavelet packet bases

Wavelet packet bases constitutes a general framework for studying dictionaries of functional bases. Indeed, they offer a large family of functional bases with several properties, depending on whether we decide to split a given subband or not [26], [27], [28], [29], among others. *Best basis* algorithms for the representation of signals involves seeking for functional atoms satisfying a given benchmark. In the wavelet framework, this benchmark is usually expressed in terms of *energy* of the coefficients, *sparsity* or number above a threshold and entropy measurements [26], [30], [31]. Recent works on *best basis* algorithms concern compressive sensing and are related to the sparsity benchmark for piecewise regular images [32].

This Section and the following Section IV provide results that complement the previous works on best basis search, in the sense that the issues addressed are: 1) the *best mutual basis* when two (or more) functions are concerned (this Section) and 2) selecting the best basis with respect to a stochasticity criterion (Section IV). We begin by addressing the mutual representation since applies for any best basis benchmark.

Assume that two different textures admit different bases for their best representation. We raise the following question: how to choose the mutual best basis with respect to the tree structures spanned by the two bases? The answer is linked on the order structure that governs the wavelet packet bases. The following provides some results on the algebraic structure of the set of wavelet packet bases. More precisely, an order relation is derived from a binary operation on the wavelet packet tree. This order relation makes basis ordering possible and eases best-mutual basis selection.

A. Order structure in wavelet packet trees

Till now, T^* denotes a full wavelet packet tree down to a fixed decomposition level J^*. Tree T^* is defined by the collection of nodes (wavelet packet subbands) $\mathsf{T}^* = \{\mathbf{W}_{j,n} : 0 \leqslant j \leqslant J^*, n = 0,1,\ldots,2^j-1 \forall j\}$. A full wavelet packet path P of T^* is defined by the sequence $\mathsf{P} = \left(\mathbf{W}_{j,n_j}\right)_{0\leqslant j\leqslant J^*}$ where $n_j = n(j) \in \{0,1,\ldots,2^j-1\}$, is recursively defined from $n_0 = 0$ and $n_\ell = 2n_{\ell-1}+\epsilon_\ell$, with $\epsilon_\ell \in \{0,1\}$ for every $1 \leqslant \ell \leqslant j$. From the above recurrence, we have that any given wavelet packet path, down to a fixed decomposition level $J \leqslant J^*$, can be completely specified by its terminal node $\mathbf{W}_{J,n_J}$ or equivalently by the binary sequence $(\epsilon_\ell)_{1\leqslant\ell\leqslant J}$ [17]. This path will hereafter be denoted by $\mathsf{P}(J,n_J) = \left(\mathbf{W}_{j,n_j}\right)_{0\leqslant j\leqslant J}$. For instance, $\mathsf{P}(0,0) = \{\mathbf{W}_{0,0}\}$ is the path consisting of the root node solely.

Consider two arbitrary wavelet packet paths $\mathsf{P}(J,n_J)$ and $\mathsf{P}(L,p_L)$. Let I be the cardinal of the set $\mathscr{A} = \{n_1,n_2,\ldots,n_J\} \cap \{p_1,p_2,\ldots,p_L\}$ and

$$q_I = \begin{cases} \max \mathscr{A} & \text{if} \quad \mathscr{A} \neq \emptyset, \\ 0 & \text{if} \quad \mathscr{A} = \emptyset. \end{cases}$$

Define an operation $\oplus$ on paths by associating to $\mathsf{P}(J,n_J) \oplus \mathsf{P}(L,p_L)$, the path $\mathsf{P}(I,q_I)$ with terminal node $\mathbf{W}_{I,q_I}$ (smallest wavelet packet space that contains $\mathbf{W}_{J,n_J}$ and $\mathbf{W}_{L,p_L}$).

A subtree T of the wavelet packet tree T^* is a collection $\mathsf{T} = \left(\mathsf{P}(J_\ell,n_{J_\ell})\right)_\ell$ of paths, where $J_\ell \leqslant J^*$ for every ℓ. Let $\mathscr{T}$ be the set of all wavelet packet subtrees of tree T^* and $\mathsf{T}_1 = \left(\mathsf{P}(J_\ell,n_{J_\ell})\right)_\ell \in \mathscr{T}$, $\mathsf{T}_2 = \left(\mathsf{P}(L_k,p_{J_k})\right)_k \in \mathscr{T}$.

Define an operation $\uplus$ on $\mathscr{T}$ by associating to $\mathsf{T}_1 \uplus \mathsf{T}_2$, the tree T_0 defined from the convention: $\mathsf{P}(I_m,q_{I_m})$ pertains to T_0 if there exists $\mathsf{P}(J_\ell,n_{J_\ell}) \in \mathsf{T}_1$ and $\mathsf{P}(L_k,p_{J_k}) \in \mathsf{T}_2$ such that $\mathsf{P}(I_m,q_{I_m}) = \mathsf{P}(J_\ell,n_{J_\ell}) \oplus \mathsf{P}(L_k,p_{J_k})$. It follows that tree T_0 is composed of nodes that are common to T_1 and T_2.

We have: $\uplus$ is a binary operation on $\mathscr{T}$ and,

Theorem 1
$(\mathscr{T},\uplus)$ is a commutative monoid with identity element T^*.

Proof: Commutativity and associativity follow from the properties of the binary operation $\oplus$ on paths. In addition, T^* is the identity element due to that any element of $\mathscr{T}$ is included in T^*. ∎

Now, let $\mathsf{P}(J,n_J)$ be a wavelet packet path. We have: $\mathsf{P}(J,n_J) \oplus \mathsf{P}(J,n_J) = \mathsf{P}(J,n_J)$. Thus, we can formulate the following proposition.

Proposition 2
The operation $\oplus$ is idempotent over the set of all wavelet packet paths.

The idempotency of $\oplus$ induces over wavelet packet paths, an order structure denoted $\preccurlyeq$ and defined by: $\mathsf{P}(J,n_J) \preccurlyeq \mathsf{P}(L,p_L) \Leftrightarrow \mathsf{P}(J,n_J) \oplus \mathsf{P}(L,p_L) = \mathsf{P}(L,p_L)$. This order relation is compatible with the operation $\oplus$. It is the inverse of the natural set ordering induced by $\subset$ operation on the wavelet packet subspaces in that the largest wavelet space with respect to set inclusion (root node $\mathbf{W}_{0,0}$) is associated with the smallest wavelet packet path.

The lower (minimal, min) and upper (maximal, max) bounds are defined by: $\mathsf{P}(J,n_J) \preccurlyeq \mathsf{P}(L,p_L) \Longleftrightarrow \mathsf{P}(J,n_J) \oplus \mathsf{P}(L,p_L) = \mathsf{P}(L,p_L) \Longleftrightarrow \min\{\mathsf{P}(J,n_J),\mathsf{P}(L,p_L)\} = \mathsf{P}(J,n_J) \Longleftrightarrow \max\{\mathsf{P}(J,n_J),\mathsf{P}(L,p_L)\} = \mathsf{P}(L,p_L)$. In addition, a subtree being a collection of paths, the collection $\mathscr{T}$ of all subtrees of T^* associated with $\uplus$ inherits the path order properties.

A full wavelet packet tree down to the decomposition level j is composed of 2^j paths whose terminal nodes are associated with frequency indices $n_j \in \{0,1,\ldots,2^j-1\}$. Consider a re-ordering of these paths obtained from a permutation G applied on the set $\{0,1,\ldots,2^j-1\}$. This permutation operates isotonically with respect to the order defined on wavelet packet paths:

$$\mathsf{P}(J,n_J) \preccurlyeq \mathsf{P}(L,p_L) \Longrightarrow \mathsf{P}(J,G(n_J)) \preccurlyeq \mathsf{P}(L,G(p_L)).$$

Thus, a re-ordering of the wavelet packet nodes such as the one involved in the Shannon wavelet packet decomposition do not impacts paths/trees ordering.

B. Infimum/supremum basis among a set of wavelet packet bases

We will say that a subtree is associated with a basis if the collection of wavelet functions generating its terminal nodes constitute a basis of the input space $\mathbf{W}_{0,0}$. This is equivalent to saying that the union of functional subspaces $\left(\mathbf{W}_{J_\ell,n_{J_\ell}}\right)_\ell$ associated with its terminal nodes equals $\mathbf{W}_{0,0}$. We seek whether an arbitrary subtree of T^* defines a basis or not. Let us define the following interval:

$$\mathscr{I}_{j,n} = \left[\frac{n}{2^j}, \frac{(n+1)}{2^j}\right[. \tag{5}$$

Then, we can formalize

Definition 2
A wavelet packet subtree $\left(\mathsf{P}(J_\ell,n_{J_\ell})\right)_\ell$ defines a basis if the intervals $\left(\mathscr{I}_{J_\ell,n_{J_\ell}}\right)_\ell$ obtained from its terminal nodes form a partition of the interval $[0,1[$.

Example 1
We have: $\left[\frac{0}{2^1},\frac{1}{2^1}\right[\cup \left[\frac{2}{2^2},\frac{3}{2^2}\right[\cup \left[\frac{3}{2^2},\frac{4}{2^2}\right[= [0,1[$. Therefore, the subtree $\left(\mathsf{P}(1,0),\mathsf{P}(2,2),\mathsf{P}(2,3)\right)$ is associated with a basis of the input space or, in other words, $\mathbf{W}_{1,0} \bigoplus \mathbf{W}_{2,2} \bigoplus \mathbf{W}_{2,3} = \mathbf{W}_{0,0}$ where $\bigoplus$ represents the direct sum between functional spaces.

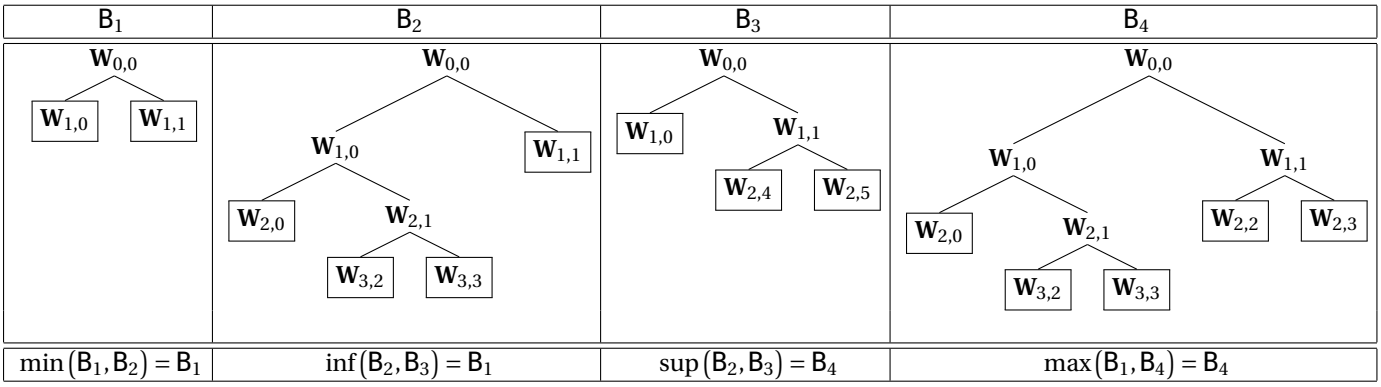

Fig. 4. Basis comparison with respect to $\preccurlyeq$.

A basis being a particular collection of subtrees, the collection $\mathscr{B}$ of all wavelet packet bases from T^* associated with $\uplus$ inherits the properties of $\mathscr{T}$:

Theorem 2
$(\mathscr{B},\uplus)$ is a commutative monoid with identity element T^*, that is the basis generated by the full wavelet packet decomposition.

The order structure on $\mathscr{B}$ is such that for two arbitrary wavelet packet bases $\mathsf{B}_1=\bigoplus_\ell \mathbf{W}_{J_\ell,n_{J_\ell}}$ and $\mathsf{B}_2=\bigoplus_k \mathbf{W}_{L_k,p_{L_k}}$ associated respectively with terminal nodes $\left(\mathsf{P}(J_\ell,n_{J_\ell})\right)_\ell$ and $\left(\mathsf{P}(L_k,p_{L_k})\right)_k$, we have: $\mathsf{B}_1\preccurlyeq\mathsf{B}_2$ if and only if every $\mathscr{I}_{J_\ell,n_{J_\ell}}$ can be written as a partition consisting in elements of $\left(\mathscr{I}_{L_k,p_{L_k}}\right)_k$. As above, this order structure makes basis comparison possible with respect to the lower (min) and upper (max) elements, as well as the greatest lower (infimum, inf) and least upper (supremum, sup) elements. Figure 4 illustrates trees and basis comparison.

Remark 1
The above order relation between wavelet packet bases/trees may have been defined only by terminal node consideration. But in practice, computing a terminal node assumes that the coefficients associated with every parent node have been computed. In addition, the basis/tree ordering obtained from terminal node consideration is not straightforward. This makes path consideration more convenient for defining the order structure given above.

Remark 2
Assume the observation of different samples from a given texture and a fixed best basis criterion (sparsity, for example). Depending on the homogeneity of the texture or the sampling procedure, the samples of the texture could admit different best bases. On one hand, the infimum and supremum trees inform us with the degree of homogeneity of the samples or the sensitivity of the sampling procedure. On the other hand, when a specific processing (similarity measurements between samples, for example) involves a fixed basis, then, we can chose either the infimum or the supremum bases, depending on their suitability.

IV. Best stochastic basis computation from the wavelet packet tree

This section provides algorithms for best wavelet packet basis selection with respect to the stochasticity criterion. Section IV-B addresses the case of a single observation and Section IV-C presents the case of several observations. These algorithms require a specified distribution model for the wavelet coefficients. In this sense, we begin by tackling in Section IV-A below, the selection of the best distribution for modeling the wavelet packet coefficients.

A. Stochastic texture retrieval using a dictionary of parametric models

The stochasticity degree obtained from $\kappa(x,F)$ in Eq. (4) assumes that x is issued from the observation of independent random variables, identically distributed, with *cdf* F. It is worth emphasizing that a single family of continuous functions, having 2 or 3 parameters, lacks enough flexibility to capture a large class of textures. In this respect, we need to build a dictionary $\mathscr{D}$ composed of numerous and appropriate continuous *cdfs*.

The literature on wavelet packet statistical properties highlights that wavelet packet coefficient distributions of many stochastic processes tend to become more regular with respect to the Gaussian distribution at large decomposition levels. This is a consequence of the decay of wavelet cumulants depending on the decomposition level and the filter order, see [17], [15] for further details. In addition, exponential decay has been proven to govern the amplitudes of the wavelet coefficients of piecewise regular functions (see [33, Sec. 6.1.3], among others). From these results, it follows that *cdfs* of *exponential family* are mostly suitable for modeling the wavelet packet coefficients. We thus consider many of such *cdfs* for constructing $\mathscr{D}$.

More precisely, the dictionary $\mathscr{D}$ used hereafter is composed of the distributions: Generalized Gaussian (GG), Weibull, Gamma, Exponential, Lognormal, Uniform and generalized extreme value. These distributions are indexed with a parameter μ. Dictionary $\mathscr{D}$ can thus be expanded as:

$$\mathscr{D}=\left\{p_{\mu,\theta},\mu\in\Upsilon,\theta=\theta(\mu)\in\Theta_\mu\right\} \tag{6}$$

where index θ refers to distribution parameters. Example: if μ =GG, then $\theta(\mu)$ =(location, scale, shape).

Let $x \equiv c_{j,n}$ be the sequence of coefficients obtained from the projection of a texture on the wavelet packet subband $\mathbf{W}_{j,n}$. The algorithm given in Table II presents the retrieval of the best distribution model for x.

TABLE II
RETRIEVAL OF THE BEST MODEL FOR x IN $\mathscr{D}$.

Compute,	for every $\mu \in \Upsilon$, $\theta_0(\mu) = \arg_{\theta\in\Theta_\mu} \max \sum_{i=1}^{N} \log f_{\mu,\theta}(x_i)$
Set:	$\mu_0 = \arg_{\mu\in\Upsilon} \min \kappa\left(x, F_{\mu,\theta_0(\mu)}\right)$ $= \arg_{\mu\in\Upsilon} \min \sup_t \left\lvert F_{x,N}(t) - F_{\mu,\theta_0(\mu)}(t)\right\rvert$.
Return:	index μ_0 and parameters $\theta_0(\mu_0)$ of the best distribution.

This algorithm provides from dictionary $\mathscr{D}$, the best distribution for representing x, that is F_{μ_0,θ_0}: the distribution pertaining to the family indexed by the argument μ_0 and whose parameters θ_0 are those obtained from the maximum likelihood over the set $\Theta(\mu_0)$ of all possible parameters $\theta(\mu_0)$ of the distribution μ_0. The stochasticity degree of subband (j,n) coefficients of the input texture is then $\kappa\left(x, F_{\mu_0,\theta_0}\right)$.

As above, T denotes a full wavelet packet tree down to a fixed decomposition level J^* and $c_{j,n} = c_{j,n}(X)$ are the coefficients of the projection of a second order random process X on $\mathbf{W}_{j,n}$. To cope with the selection of the best wavelet filter, we assume that the wavelet order is fixed to 7: Daubechies wavelet with 7 vanishing moments is used. This choice is motivated by [18], [16], [15] where this order is proven to be reasonable for decorrelating a large class of random processes.

B. Best stochastic basis for the representation of a texture

The algorithm below makes it possible to find, for a given texture and a fixed stochasticity bound, the best stochastic wavelet packet basis with respect to $\mathscr{D}$. The Best Stochastic Basis (BSB) research in Wavelet Packet (WP) tree (BSB-WP) is given in Table III. This algorithm assumes that the maximum depth of the wavelet packet tree is fixed to J^*.

BSB-WP provides the best basis (subtree of T) in the sense that a stochastic node is no more decomposed: the coefficients issued from this node have already reached the stochasticity degree imposed by η so that decomposing this node is not appropriate. As consequence, the texture can be represented with *precison* η by a sequence of elements of $\mathscr{D}$ applied for modeling subbands of the BSB-WP tree.

TABLE III
BSB-WP ALGORITHM: BEST STOCHASTIC BASIS SEARCH IN A WAVELET PACKET TREE WITH MAXIMUM DEPTH FIXED TO J^*.

Do
  Decompose any non-stochastic node $c_{j,n}$.
  Check[4] the stochasticity of $(c_{j+1,2n+\epsilon})_{\epsilon=0,1}$ with respect to the distribution models given in $\mathscr{D}$ and the bound η.
Until All nodes are stochastic and $j \leqslant J^*$.
Retrieve the nodes associated with the stochastic subbands.

Remark 3
Since, in general, the approximations of non-stationary processes are not expected to become stationary [15], then, when the remaining non-stochastic subband is the coarser approximation subband $(J^*,0)$, we will consider the input texture as the sum of a deterministic pattern represented by the approximation contribution and a stochastic process resulting from the contribution of the BSB-WP detail nodes.

Remark 4
If there exist remaining non-stochastic detail subbands at decomposition level J^*, then we will say that the texture admits no stochastic representation with respect to bound η by using a subtree of T and dictionary $\mathscr{D}$.

C. Which basis for mutual representation or dual analysis?

TABLE IV
COMPUTATION OF THE SUPREMUM OF STOCHASTIC BASES.

Set *node unions* be an empty set
For every stochastic basis:
  Get all nodes associated with the paths having terminal nodes, the nodes of the stochastic basis.
  Compute the new *node unions* as the union between these nodes and the nodes obtained at the former iteration.
End
Retrieve the terminal nodes associated with sequence *node unions*. These terminal nodes define the supremum basis among the bases considered.

TABLE V
COMPUTATION OF THE INFIMUM OF STOCHASTIC BASES.

Let J^* the larger decomposition level involved in the bases.
Set *node intersections* be the sequence composed with all nodes involved in a full wavelet packet tree down to level J^*.
For every stochastic basis:
  Get all nodes associated with the paths having terminal nodes, the nodes of the stochastic basis.
  Compute the new *node intersections* as the intersection between these nodes and the nodes obtained at the former iteration.
end
Retrieve the terminal nodes associated with sequence *node intersections*. These terminal nodes define the infimum basis among the bases considered.

This section provides algorithms for finding the stochastic bases that encompass a given set of best bases for the representation of a family of textures. Without loss of generality, we assume that two textures are concerned and that they yield two different stochastic bases at degree η_i for their representation. The outstanding bases are the infimum and supremum bases. Indeed, the infimum and supremum bases (see Section III-B) are the closest bases which inherit approximately the same properties as the two given bases. For instance, the supremum stochastic basis is the smallest basis that yields joint stochastic representation. Now, when discriminant analysis has to be performed on textures, then it is reasonable to measure similarity on the infimum basis:

decomposing beyond the terminal nodes of this basis is not expected to yield better representation for the textures. However, note that, beyond this algebraic consideration, selection of a criterion for the joint representation of these textures still depends on the problem tackled. The algorithms for finding the joint optimal bases are detailed in Tables IV and V.

V. Application

A. Texture classification

TABLE VI
VisTeX "Fabric" texture classification from BSB-WP stochasticity measurements. The "Fabric" textures are given in figure 5. In this table, "Det" designate wavelet packet details and "Approx" designate wavelet approximations.

	Quasi-stochastic		**Stochastic**		**Strongly-stochastic**	
	Det.	Approx.	Det.	Approx.	Det.	Approx.
Fabric.18	√	√	√	√	√	–
Fabric.07	√	√	√	√	–	–
Fabric.17	√	√	√	–	–	–
Fabric.04	√	√	√	–	–	–
Fabric.09	√	√	√	–	–	–
Fabric.11	√	–	√	–	–	–
Fabric.15	√	–	√	–	–	–
Fabric.00	–	–	–	–	–	–
Fabric.14	–	–	–	–	–	–

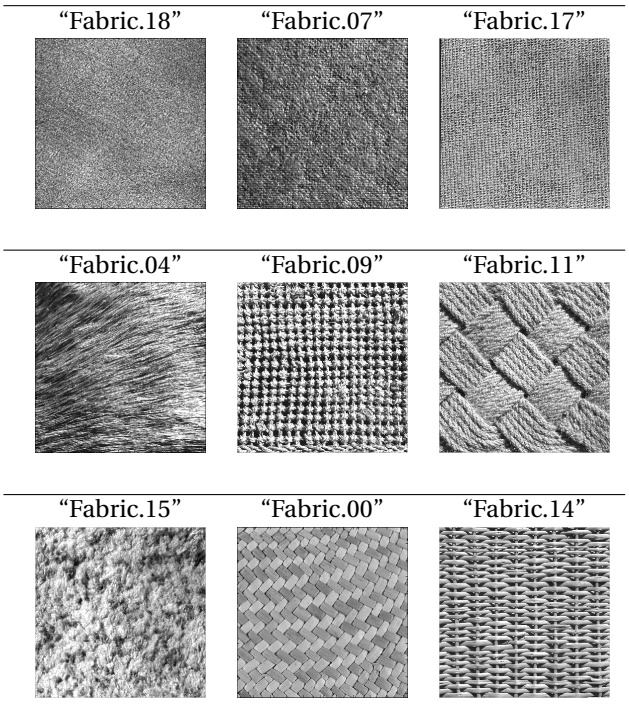

Fig. 5. Textures "Fabrics" from the VisTeX database.

We run the BSB-WP algorithm in order to classify the subclass of "Fabric" images of the VisTeX textures, with respect to the semantic 4 classes (non-stochastic, quasi-stochastic, stochastic, strongly-stochastic) given by Definition 1. We use a maximum depth J^* fixed to 4 for the wavelet packet decomposition and the Daubechies wavelet of order 7 is used. Table VI summarizes the results obtained, the corresponding textures being given in Figure 5.

The comparison between the results of this table and the visual perception of the corresponding textures highlights that the randomness appearance measured from parameter κ is coherent with the natural perception of stochasticity. Some textures ("Fabric.18", "Fabric.07") can be well represented by a sequence of parametric models applied on all subbands of the best basis. Non stochastic textures ("Fabric.00", "Fabric.14") are such that WP parametric modeling is not relevant for their characterization because WP coefficient distribution (that tend to be sparse because these textures are piecewise regular) deviate significantly from the continuous *cdfs* composing dictionary $\mathscr{D}$.

The following provides some examples for illustrating BSB-WP texture characterization.

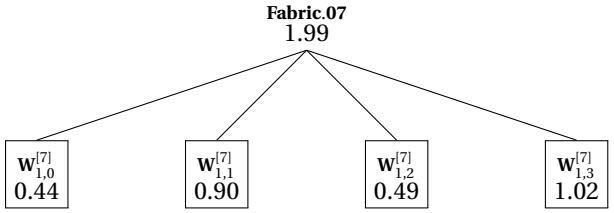

Fig. 6. $100\times\kappa$ for texture "Fabric.07" from the VisTeX database. The BSB-WP is composed of framed subbands.

Example 2
Texture "Fabric.07" is intrinsically quasi-stochastic: $\kappa_{\text{Fabric.07}} < \eta_0$. BSB-WP provides a basis (see the basis composed of framed subbands in Figure 6) where all subbands involved in the representation are stochastic: $\kappa_{c_{1,n}[\text{Fabric.07}]} < \eta_1$ for every $n = 0,1,2,3$. BSB-WP basis with higher stochasticity property ($\kappa < \eta_2$) has not been found up to decomposition level J^*.

Example 3
Texture "Fabric.09" is not intrinsically stochastic: $\kappa_{\text{Fabric.09}} > \eta_0$. BSB-WP provides a basis where the texture can represented as a deterministic (smooth) approximation and stochastic details: $\kappa_{c_{j,n}[\text{Fabric.09}]} < \eta_1$ for every nodes (j,n) involved in the tree of Figure 7, with $n \neq 0$.

Example 4
Texture "Fabric.14" is not stochastic. κ-measurements are out of stochasticity bounds for the input texture as well as for many of its wavelet packet subbands up to decomposition level J^*. In addition, this texture presents a "singular" path in the sense given in [15]. Indeed, depending on the input process, some paths are such that no *cdf* regularization can be expected. In this case, stochasticity measures can increase as the decomposition level increases. This occurs for the path with subbands marked in oval frames in Figure 8.

It follows that wavelet based stochasticity measurements reflect the "randomness-like contribution" or "randomness degree" of textures. This randomness degree reflects the

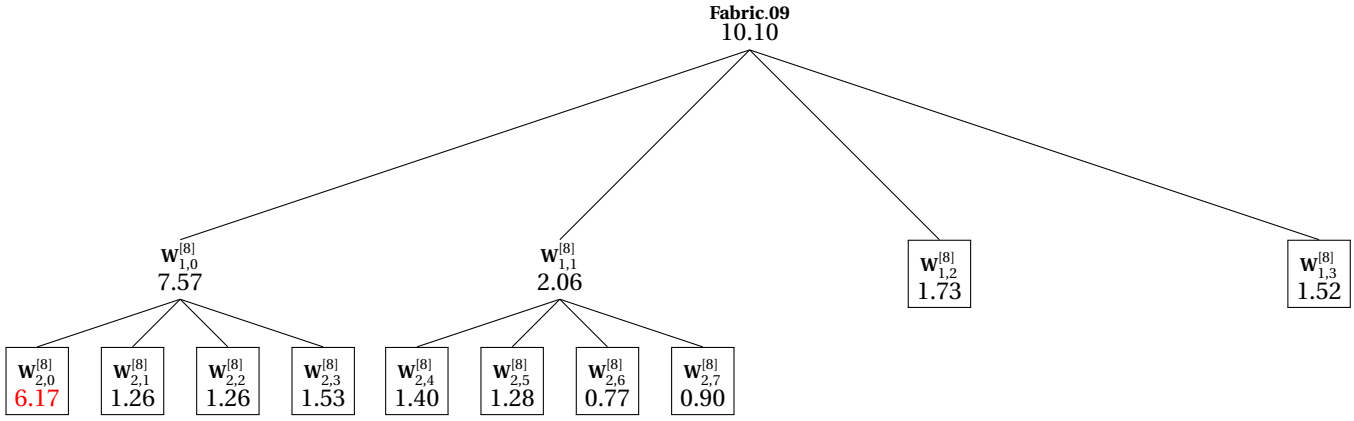

Fig. 7. $100\times\kappa$ for texture "Fabric.09" from the VisTeX database. The BSB-WP is composed of framed subbands. The texture is represented as the sum of a smooth approximation and stochastic details.

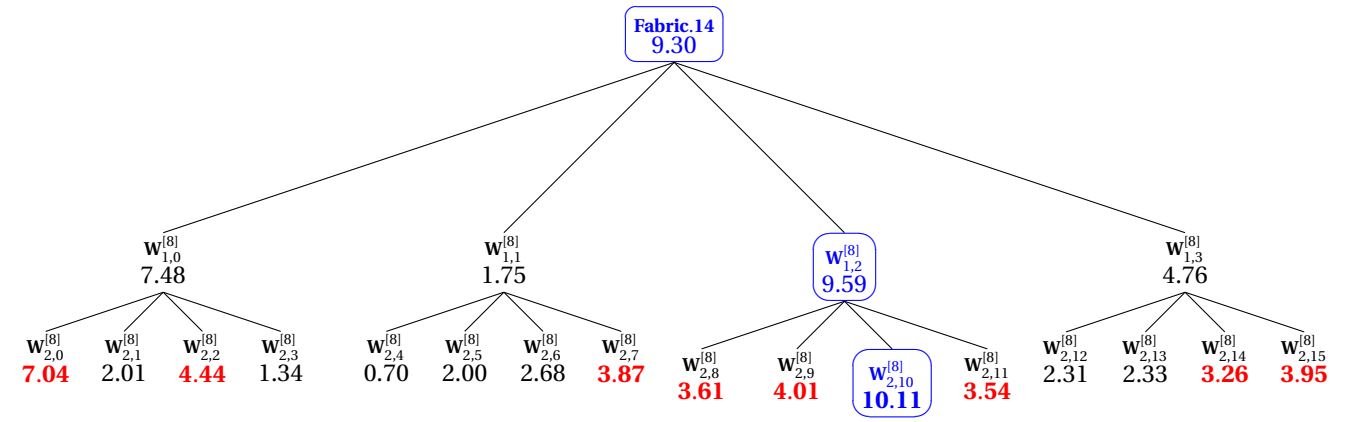

Fig. 8. $100\times\kappa$ for texture "Fabric.14" from the VisTeX database at decomposition level 2. Many subbands remain non-stochastic. In addition, the stochasticity parameter κ does not systematically decrease as the decomposition level increases in some detail paths (see the path with oval boxes).

natural perception of textures (see Figure 5 and Table VI for illustration) in that the human vision is known to be sensitive to the roughness/coarseness/coherence of observations. For instance, when SAR image browsing is of interest, the query of a farm from that of a forest mainly differs from the roughness that characterizes these elements. The same holds true when looking for a particular material in geosciences: the difference of perception between rocks (resp. geological strata) consists mainly in the intrinsic coarseness of their particles (resp. layers). The following provides two methods for content-based texture retrieval by using stochasticity measurements.

B. Content-based stochasticity retrieval

This section addresses stochasticity consideration in Content-Based Image Retrieval (CBIR) in texture databases. Parametric modeling of the wavelet coefficients is known to be efficient in CBIR (see [34], [35], [36], [37], [38] and [39], for instance). The CBIR under consideration assumes that the model distribution of the query and those of the elements in the database are identified. This corresponds to be the level 1 CBIR [40]: the query, an image, is completely specified through its pixel values. The parametric modeling associated with these values is used to describe the statistical distribution of the pixels (low-level feature description).

Consider now the level 2 CBIR [40], most commonly used in practical applications, where the query is specified by some of its high-level features. The main question concerns feature selection, features that make inference possible with respect to the human perception. For structured objects with geometrically regular edges, the form (through primitives) or the regularity of the objects can serve as features and a query object can be inferred through these features, see [10], [41].

In absence of structured components that can be taken as references, then the randomness degree measured by the stochasticity parameter can be used to infer the intrinsic coherence (non-coherence being close to stochasticity) of the observation and thus, pass by the model. In this sense, we deal with level 2 Content-Based Coherence/Stochasticity Retrieval (CBCR/CBSR).

The main contribution of the present paper with respect to CBIR is the breaking of a semantic gap through texture CBSR. This is performed by: 1) relating parametric modeling of textures to the concept of randomness degree, 2) learning the tree structure that makes stochastic representation of a given family of texture possible at degree η (example: the query "forest" will be associated to a specific tree representation delimited by the infimum and supremum trees) and 3) learning the stochasticity degree of subband wavelet

coefficients, given a fixed tree, in order to characterize a texture through the variation of its stochasticity parameters.

In the following, we provide details on the level 2 CBSR approaches proposed. We first describe in Section V-B1, how to learn the tree structure that yields stochastic representation at a fixed stochasticity degree η. An illustrative example is given for some "Fabric" textures. Then, Section V-B2 provides CBRS by learning the stochasticity degrees for a fixed tree, with detailed experiments based on VisTeX and Brodatz databases.

1) Content-based stochasticity retrieval by learning the stochasticity tree structure: In this section, we consider a set of M texture classes indexed by integer $m, 1 \leqslant m \leqslant M$. For a given class m, we assume that samples $(m_k)_{k=1,2,\ldots K_m}$ are avalaible (learning database). Let $\mathsf{B}_{\text{Best}}[m_k]$ denotes the best basis associated with sample m_k at the fixed stochasticity degree η.

Consider the infimum and supremum of the best stochasticity bases (see Section III) associated with samples $(m_k)_k$ of texture class m and denoted by

$$\mathsf{B}^{\inf}[m] = \inf\left\{\mathsf{B}_{\text{Best}}[m_1], \mathsf{B}_{\text{Best}}[m_2], \ldots, \mathsf{B}_{\text{Best}}[m_{K_m}]\right\},$$
$$\mathsf{B}^{\sup}[m] = \sup\left\{\mathsf{B}_{\text{Best}}[m_1], \mathsf{B}_{\text{Best}}[m_2], \ldots, \mathsf{B}_{\text{Best}}[m_{K_m}]\right\}.$$

The tree structure describing the behaviour of the best stochastic representations of the samples of texture m have lower bound $\mathsf{B}^{\inf}[m]$ and upper bound $\mathsf{B}^{\sup}[m]$.

The CBSR principle considered in this section is the following: an arbitrary sample belongs to stochasticity class m if its best basis at stochasticity degree η, denoted by B, is such that: $\mathsf{B}^{\inf}[m] \preceq \mathsf{B} \preceq \mathsf{B}^{\sup}[m]$.

Consider the set of "Fabric" textures from the VisTeX database (see Figure 5). From the classification obtained in Table VI, we focuses on "Fabric.0007" and "Fabric.0018" which are closer on the basis of their stochasticity degrees. We set the stochasticity degree to η_2. We then consider the following experimental setup: each image is splited into 16 non-overlapping subimages, $K = 8$ images among them (the 8 upper-half subimages) are used as the training set. The remaining 16 subimages, 8 subimages of 'Fabric.0007" and 8 subimages of "Fabric.0018", the lower-half subimages, are put together to form the test database.

We run the following CBSR strategy:

- *Learn* the stochasticity tree structure for any of the "Fabric" texture by computing $\mathsf{B}^{\inf}$ and $\mathsf{B}^{\sup}$ from its 8 samples available from the learning database.
- *Retrieve*, from the test database, the samples that belong to the semantic class of any of the "Fabric" texture, that are the samples having stochasticity bases bounded by the infimum and supremum bases associated with the class.
- *Sort* the samples thus obtained and compute texture-specific retrieval.

From the experiments carried out, we have that:

- The learned basis structure corresponding to "Fabric.0007" is any basis B such that:

$$\bigcup_{n=0,1,2,3} \mathbf{W}_{1,n} \preceq \mathsf{B} \preceq \mathbf{W}_{2,0} \cup \left(\bigcup_{n=4,5,\ldots,4^3-1} \mathbf{W}_{3,n} \right).$$

- The learned basis structure for "Fabric.0018" is any basis B such that

$$\mathbf{W}_{2,0} \cup \left(\bigcup_{j=1,2} \bigcup_{n=1,2,3} \mathbf{W}_{j,n} \right) \preceq \mathsf{B} \preceq \mathbf{W}_{3,0} \cup \left(\bigcup_{j=1,2,3} \bigcup_{n=1,2,3} \mathbf{W}_{j,n} \right).$$

The retrieval rates obtained from the test database are such that:

- Query "Fabric.0007", associated with the lowest randomness degree among the two classes, reduces the search database from 16 to 8 including 7 good retrieval/8.
- Query "Fabric.0018", associated with the highest randomness degree among the two classes, reduces the search database from 16 to 7 including 7 good retrieval/8.

From these experiments, it follows that we can characterize the statistical distributions of the coefficients of a texture (low-level features) by a tree structure with respect to stochasticity bounds and perform high-level CBSR with query concepts consisting in higher or lower randomness appearance.

2) Content-based stochasticity retrieval by learning the stochasticity bounds: Depending on constraints such as computational load or dealing with a large number of semantic classes, it may sometimes be desirable to fix the decomposition basis. In this section, we consider a fixed wavelet packet basis $\mathsf{B} = \bigcup_{p=1,2,\ldots,L} \mathbf{W}_{J_p,n_p}$ and propose high-level CBSR by learning the subspace (hypercube) where the stochasticity parameters are expected to lie within.

Assume the availability of K samples (subimages) for any texture class m considered, with $1 \leqslant m \leqslant M$ (training set for this texture). Let us denote by $\kappa_{m_\ell}(J_p, n_p)$, the value of the stochasticity parameter (see Eq. (4)) associated with the subband $\mathbf{W}_{J_p,n_p}$ coefficients of subimage m_ℓ. The sequence $\left(\kappa_{m_\ell}(J_p, n_p)\right)_{\ell=1,2,\ldots,K}$ represents the behaviour of the stochasticity parameters of the projection of texture m samples on subband $\mathbf{W}_{J_p,n_p}$. Let us denote $\kappa^m_{\min}(J_p, n_p) = \min\left\{\kappa_{m_\ell}(J_p, n_p), \ell = 1, 2, \ldots, K\right\}$ and $\kappa^m_{\max}(J_p, n_p) = \max\left\{\kappa_{m_\ell}(J_p, n_p), \ell = 1, 2, \ldots, K\right\}$. Define the stochasticity hypercube associated with the samples of texture m on basis B by

$$\mathscr{H}^m_L = \prod_{p=1}^{L} \left[\kappa^{\min}_m(J_p, n_p), \kappa^{\max}_m(J_p, n_p)\right].$$

The CBSR principle considered in this section is the following: a query sample admitting stochasticity parameters $\kappa(J_1, n_1), \kappa(J_2, n_2), \ldots, \kappa(J_L, n_L)$ on basis B is decided to belong to class m if the vector $(\kappa(J_1, n_1), \kappa(J_2, n_2), \ldots, \kappa(J_L, n_L)) \in \mathscr{H}^m_L$.

In this respect, a texture can be characterized by the hypercube defined from the lower and upper bounds of the stochasticity parameters of its sample coefficients on the basis B. This hypercube defines the semantic class of the texture.

Assume that a new sample of the texture is available. Then we can re-evaluate the stochasticity bounds when some of the additional stochasticity parameters of this

TABLE VII
TRUE POSITIVE RATE (TPR) AND FALSE ALARM RATE (FAR) FOR CBRS BY LEARNING THE STOCHASTICITY BOUNDS. DICTIONARY $\mathscr{D}$ IS USED FOR STOCHASTICITY MEASUREMENTS.

Texture	TPR	FAR	Texture	TPR	FAR
Bark.00	62.50	08.97	Food.08	56.25	0
Bark.06	37.50	07.69	Grass.01	37.50	04.81
Bark.08	62.50	00.80	Leav.08	68.75	12.66
Bark.09	75.00	10.58	Leav.10	43.75	09.13
Bric.01	31.25	01.44	Leav.11	56.25	04.65
Bric.04	37.50	02.56	Leav.12	43.75	04.65
Bric.05	50.00	06.89	Leav.16	62.50	01.92
Buil.09	37.50	00.96	Meta.00	50.00	01.92
Fabr.00	43.75	01.92	Meta.02	68.75	00.32
Fabr.04	50.00	07.05	Misc.02	62.50	00.64
Fabr.07	62.50	01.12	Sand.00	31.25	02.24
Fabr.09	37.50	00.48	Ston.01	56.25	08.01
Fabr.11	56.25	01.44	Ston.04	62.50	01.92
Fabr.14	50.00	0	Terr.10	50.00	08.01
Fabr.15	68.75	00.80	Tile.01	31.25	02.40
Fabr.17	56.25	01.12	Tile.04	37.50	00.96
Fabr.18	68.75	00.48	Tile.07	25.00	0
Flow.05	31.25	07.21	Wate.05	62.50	02.40
Food.00	62.50	01.92	Wood.01	56.25	12.18
Food.05	31.25	03.21	Wood.02	56.25	09.29

TABLE VIII
TRUE POSITIVE RATE (TPR) AND FALSE ALARM RATE (FAR) FOR CBRS BY LEARNING THE STOCHASTICITY BOUNDS. THE GG FAMILY IS USED FOR STOCHASTICITY MEASUREMENTS.

Texture	TPR	FAR	Texture	TPR	FAR
Bark.00	100	18.27	Food.08	75.00	00.48
Bark.06	75.00	13.78	Grass.01	50.00	07.85
Bark.08	68.75	03.37	Leav.08	75.00	16.03
Bark.09	75.00	20.83	Leav.10	56.25	14.58
Bric.01	62.50	03.85	Leav.11	62.50	06.89
Bric.04	68.75	04.01	Leav.12	43.75	04.81
Bric.05	75.00	08.97	Leav.16	75.00	02.56
Buil.09	56.25	02.56	Meta.00	68.75	02.72
Fabr.00	62.50	02.88	Meta.02	87.50	00.64
Fabr.04	50.00	13.62	Misc.02	68.75	00.96
Fabr.07	81.25	01.44	Sand.00	62.50	03.37
Fabr.09	75.00	00.64	Ston.01	62.50	13.94
Fabr.11	81.25	02.40	Ston.04	68.75	03.53
Fabr.14	87.50	00.16	Terr.10	68.75	16.35
Fabr.15	87.50	02.08	Tile.01	37.50	04.17
Fabr.17	87.50	03.05	Tile.04	68.75	02.08
Fabr.18	81.25	01.12	Tile.07	31.25	0
Flow.05	56.25	11.06	Wate.05	81.25	07.69
Food.00	93.75	02.56	Wood.01	93.75	24.20
Food.05	50.00	04.17	Wood.02	75.00	15.22

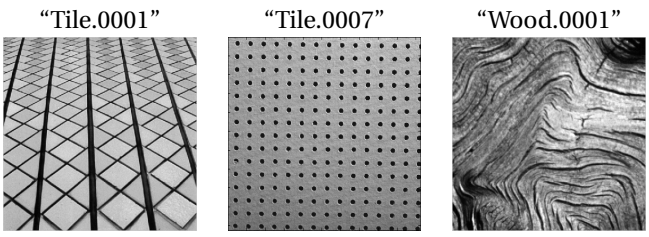

Fig. 9. Textures "Tile.0001", "Tile.0007", "Wood.0001" from the VisTeX album.

sample are out of the texture stochasticity hypercube. In addition, depending on the distribution of the stochasticity parameters, the user can discard those behaving as outliers in order to tighten the stochasticity hypercube and avoid overlapping with stochasticity hypercubes associated to other semantic classes. This re-evaluation is known to be useful in integrated CBIR systems [42].

The following provides CBSR experimental results obtained for $M = 40$ textures from the VisTeX database. The experimental setup used is described below:

- First, we construct the learning database by using the top-half of the images: each top-half image is splited into $K = 8$ non-overlapping subimages (128 × 128 pixels per subimage). These K subimages are used to compute the stochasticity hypercube $\mathscr{H}_L^m$ for $m = 1, 2, \ldots, 40$.
- Then, we constitute the test database by using the down-half of the images: each down-half image is splited into 8 non-overlapping subimages. Thus, the test database is composed of $8 \times M$ subimages.
- In order to increase the number of experiments, we have also permutated the roles played by the learning and the test database (top-half becomes down-half and *vice-versa*).

We run this procedure when the decomposition is performed by using a wavelet basis with $J^* = 2$. The stochasticity is measured with respect to dictionary $\mathscr{D}$ (Table VII) and with respect to a single distribution family: the GG distributions (Table VIII). Specifically, in these tables, we have that 2 stochasticity coordinates out of $\mathscr{H}_L^m$ are tolerated, that is, 2 stochasticity parameters out-of-bounds are tolerated among a set of $3 * J^* + 1 = 7$ stochasticity parameters $\kappa(J_1, n_1), \kappa(J_2, n_2), \ldots, \kappa(J_L, n_L)$ with $L = 7$. In Tables VII and VIII, TPR denotes the True Positive Rate defined as the ratio (fraction relevant queries per class):

$$\mathrm{TPR}[m] = \frac{\text{Number of admissible subimages that are issued from texture } m}{\text{Total number of relevant subimages,}}$$

and FAR denotes the False Alarm Rate per class:

$$\mathrm{FAR}[m] = \frac{\text{Number of admissible subimages that are not issued from texture } m}{\text{Total number of subimages that are not issued from texture } m}$$

As it can be seen in these tables, stochasticity based retrieval is relevant for most textures given in this database. Low TPRs occur when texture is very regular (Example: "Tile.0001", "Tile.0007"), see Figure 9. High FARs occur when texture have non-homogeneous subimages (Example: "Wood.0001"): the bounds define a large interval which is expected to contain stochasticity values related to many other textures, see Figure 9.

Experiments on the Brodatz album yield approximately the same results. The global TPR is 69% for the Brodatz album (resp. 70% for the VisTeX album) and the global FAR is 10% for the Brodatz album (resp. 7% for the VisTeX album), when GG modeling is used for stochasticity measurements. Tables concerning Brodatz album are omitted because 111 textures are concerned by the tests.

VI. CONCLUSION

The paper has provided algorithmic tools for best basis selection with respect to stochasticity criterion. When regularity fail to be relevant because measuring low regularity is intricate, stochasticity can be used efficiently.

The framework used for measuring stochasticity is that of the dictionary of wavelet packet bases because of their suitable algebraic and statistical properties. The order structure involved in wavelet packet bases makes it possible to define the infimum and supremum bases among a set a best stochastic bases. The relevance of the stochasticity analysis through wavelet packet tree has proven efficiency in classification and content-based stochasticity retrieval. The best stochastic basis selection can be extended straightforwardly to larger dictionaries of functional bases or when considering union of several dictionaries with the same order structure. Further prospects concerning this work are related to replacing the stochasticity hypercubes by smooth manifolds in order to limit miss-classifications. Finally, a question that arises is: up to what extent can we guess a duality between the notions stochasticity and regularity?

REFERENCES

[1] A. P. Korostelev and A. B. Tsybakov, *Minimax theory of image reconstruction*. Springer-Verlag, New York, 1993.

[2] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, pp. 460 – 473, 1978.

[3] B. Vasselle and G. Giraudon, "A multiscale regularity measure as a geometric criterion for image segmentation," *Machine Vision and Applications*, vol. 7, no. 4, pp. 229 – 236, 1994.

[4] B. S. Manjunath, J. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 703 – 715, 1998.

[5] D. Chetverikov, "Pattern regularity as a visual key," *Image and Vision Computing*, vol. 18, pp. 975 – 985, 2000.

[6] K. Fujii, S. Sugi, and Y. Ando, "Textural properties corresponding to visual perception based on the correlation mechanism in the visual system," *Psychological Research*, vol. 67, no. 3, pp. 197 – 208, 2003.

[7] L. Trujillo, G. Olague, P. Legrand, and E. Lutton, "Regularity based descriptor computed from local image oscillations," *Optics Express*, vol. 15, no. 10, pp. 6140 – 6145, 2007.

[8] J. G. Daugmann, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America. A, Optics and Image Science*, vol. 2, no. 7, pp. 1160 – 1169, 1985.

[9] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 722 – 733, 1996.

[10] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947 – 963, 2001.

[11] Y. Stitou, F. Turcu, Y. Berthoumieu, and M. Najim, "Three-dimensional textured image blocks model based on wold decomposition," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3247 – 3261, 2007.

[12] A. N. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *G. Ist. Ital. Attuari*, vol. 4, pp. 83 – 91, 1933.

[13] V. Arnold, "Orbits' statistics in chaotic dynamical systems," *Nonlinearity*, vol. 21, pp. T109 – T112, 2008.

[14] V. G. Gurzadyan and A. A. Kocharyan, "Kolmogorov stochasticity parameter measuring the randomness in the cosmic microwave background," *Astronomy & Astrophysics*, vol. 492, no. 2, pp. L33 – L34, Dec. 2008.

[15] A. M. Atto and Y. Berthoumieu, "Wavelet transforms of non-stationary random processes: Contributing factors for stationarity and decorrelation," *Preprint, [Available Online]*, Jan. 2011.

[16] A. M. Atto, D. Pastor, and G. Mercier, "Wavelet packets of fractional brownian motion: Asymptotic analysis and spectrum estimation," *IEEE Transactions on Information Theory*, vol. 56, no. 9, Sep. 2010.

[17] A. M. Atto and D. Pastor, "Central limit theorems for wavelet packet decompositions of stationary random processes," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 896 – 901, Feb. 2010.

[18] A. M. Atto, D. Pastor, and A. Isar, "On the statistical decorrelation of the wavelet packet coefficients of a band-limited wide-sense stationary random process," *Signal Processing*, vol. 87, no. 10, pp. 2320 – 2335, Oct. 2007.

[19] J. F. J. Massey, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 253, no. 2, pp. 68 – 78, Mar. 1951.

[20] J. F. Kenney and E. S. Keeping, *Mathematics of Statistics, Pt. 2*, Princeton, Ed. Van Nostrand, 1951.

[21] P. Flandrin, "Wavelet analysis and synthesis of fractional brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, Mar. 1992.

[22] J. Zhang and G. Walter, "A wavelet-based KL-like expansion for wide-sense stationary random processes," *IEEE Transactions on Signal Processing*, vol. 42, no. 7, pp. 1737–1745, July 1994.

[23] T. Kato and E. Masry, "On the spectral density of the wavelet transform of fractional brownian motion," *Journal of Time Series Analysis*, vol. 20, no. 50, pp. 559–563, 1999.

[24] D. Leporini and J.-C. Pesquet, "High-order wavelet packets and cumulant field analysis," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 863–877, Apr. 1999.

[25] P. F. Craigmile and D. B. Percival, "Asymptotic decorrelation of between-scale wavelet coefficients," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 1039 – 1048, Mar. 2005.

[26] M. V. Wickerhauser, "Inria lectures on wavelet packet algorithms," *in* Problèmes Non-Linéaires Appliqués, Ondelettes et Paquets D'Ondes, *P.-L. Lions, Ed., INRIA, Roquencourt, France*, pp. 31 – 99, June 1991.

[27] I. Daubechies, *Ten lectures on wavelets*. SIAM, Philadelphie, PA, 1992.

[28] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713 – 718, Mar. 1992.

[29] N. Hess-Nielsen and M. V. Wickerhauser, "Wavelets and time-frequency analysis," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 523 – 540, Mar. 1996.

[30] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1186 – 1191, 1993.

[31] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Transactions on Image Processing*, vol. 2, no. 4, pp. 429 – 441, 1993.

[32] G. Peyré, "Best basis compressed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2613 – 2622, May. 2010.

[33] S. Mallat, *A wavelet tour of signal processing, second edition*. Academic Press, 1999.

[34] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146 – 158, Feb. 2002.

[35] R. Kwitt and A. Uhl, "Image similarity measurement by kullback-leibler divergences between complex wavelet subband statistics for texture retrieval," *IEEE International Conference on Image Processing, ICIP*, San Diego, California, USA, 12-15 October, pp. 933–936, 2008.

[36] N. Lasmar, Y. Stitou, S. Jouini, Y. Berthoumieu, and M. Najim, "Parametric gaussianization procedure of wavelet coefficients for texture retrieval," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Las Vegas, Nevada, USA, 30 march - 4 April, pp. 749 – 752, 2008.

[37] M.-C. Peron, J.-P. D. Costa, Y. Stitou, C. Germain, and Y. Berthoumieu, "Joint linear-circular stochastic models for texture classification," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Las Vegas, Nevada, USA, 19 - 24 April, pp. 1073 – 1076, 2009.

[38] Y. Stitou, N. Lasmar, and Y. Berthoumieu, "Copulas based multivariate gamma modeling for texture classification," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Las Vegas, Nevada, USA, 19 - 24 April, pp. 1045 – 1048, 2009.

[39] N. Lasmar and Y. Berthoumieu, "Multivariate statistical modeling for texture analysis using wavelet transforms," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Dallas, Texas, USA, 14 - 19 March, 2010.

[40] J. Eakins and M. Graham, "Content-based image retrieval," *University of Northumbria at Newcastle*, Technical Report, 1999.

[41] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262 – 282, 2007.

[42] A. W. M. Smeulders, M. Worring, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349 – 1380, 2000.