DWT-BASED NON-PARAMETRIC TEXTURE MODELING

G. Menegaz

Audiovisual Communications Laboratory
Swiss Federal Institute of Technology,
CH–1015 Lausanne
e-mail: Gloria.Menegaz@epfl.ch

ABSTRACT

We propose a non-parametric texture modeling and synthesis technique based on the integer version of the Discrete Wavelet Transform (DWT). The successive levels of the DWT pyramid of the input texture are progressively sampled starting from the signal approximation to generate the analogous wavelet pyramid for the synthetic texture. An underlying statistical model is assumed, where the appearance of wavelet coefficients at each scale is conditioned by the appearance of the corresponding *ancestors* at coarser scales. A non-parametric Parzen estimator is used for sampling. The integer DWT is obtained by the lifting steps implementation. The proposed method provides results comparable to the other state-of-the-art techniques for random (unstructured) textures, but at a very low computational complexity. For structured textures, performance depends on the specific orientation features and structure size.

1. INTRODUCTION

Image modeling is the core of perceptual image processing. The potential of this bridging discipline towards physiology is enormous. The ability to describe an image in terms of those parameters that are relevant for visual perception would not only support a great number of applications in the fields of telecommunications and multimedia, but also provide feedback to physiological investigation. Neurophysiological evidence has shown that the Human Visual System (HVS) performs a multiresolution analysis of the visual stimuli. The receptive fields of simple cells are localized in space and time, have band-pass characteristics in the spatial and temporal frequency domains, are oriented, and are often sensitive to the direction of motion of a stimulus [1, 2]. Several probabilistic texture models have grown on the insights coming from neural sciences. Among the most relevant are the ones proposed by Portilla and Simoncelli [3], Zhu [4], and De Bonet [5, 6].

In this paper we aim at investigating the exploitability of a simple and low complexity subband decomposition the integer DWT, in the framework of perception-based texture modeling. Even though such critically sampled transformation selecting only three fixed orientations is not suitable for mimicking the HVS, it has proven to be well representative of some texture classes, and thus very useful for applications aiming at the *visually lossless* representation of a given reference texture. A typical example is model-based coding, where the goal is to increase the compression efficiency by replacing the "real" information by some synthetic version producing the same visual effect [7].

Our method is inspired by the non-parametric multi-scale statistical model for natural images presented in [5, 6]. What we have retained of this method is the high-level procedure, meaning the idea of using the original texture as *model* and directly *sample* its multiresolution features to progressively *build* a *perceptually equivalent* multiscale representation of the synthetic texture. The multiresolution sampling procedure is such that each spatial frequency band of the synthetic image is obtained by *sampling*, according to a predefined criterion, the corresponding spatial frequency bands of the original texture. The criterion basically consists in conditioning the sampling process on the occurrence of local features at coarser resolutions. Our approach is different from De-Bonet's Multiresolution Probabilistic Texture Modeling (MPTM) [5] in many respects. First, it is entirely based on the integer DWT transform [8]. This makes it particularly suitable for the integration in a DWT-based coding system. No additional processing would be required for the definition of the texture model, which could then be *classically* encoded. Furthermore, the lifting steps scheme asymptotically reduces the computational complexity by a factor of 4 [9], and all the transform process can be performed with integer arithmetic, which is quite advantageous for the implementation on a device. Second, one single pyramid is used for both conditioning and sampling. We call it *hyper-pyramid* (HP) to emphasize that the three orientation subbands of each DWT level are representative of one level of the *analysis&conditioning* pyramid.

This paper is organized as follows. Section 2 gives an overview on MPTM approach. The proposed DWT-based method is presented in Sec. 3. Results are discussed in Sec. 4, and conclusions are derived in Sec. 5.

2. MULTIRESOLUTION PROBABILISTIC TEXTURE MODELING (MPTM)

The goal of probabilistic texture synthesis is to generate a new image from a sample textures which is *sufficiently different* from the original, yet appears to be generated by the *same underlying stochastic process*. Basically, a good compromise must be reached between a certain degree of *randomization*, which makes the synthetic texture distinguishable from the original one, and the preservation of the features which determine the *visual appearance* of the texture. The solution proposed in [5, 6] simplifies this task by synthesizing textures which look similar at low spatial resolution, and then maintaining that similarity while progressing towards higher frequencies. Accordingly, in the MPTM approach the input image is first analyzed by measuring the joint occurrence of texture discrimination features at different scales. Then, a new

0-7803-6725-1/01/$10.00 ©2001 IEEE

texture is generated by *sampling* successive frequency bands from the input texture, conditioned on the similar joint occurrence of features at lower spatial frequencies. In [5, 6], the pyramid of the original texture is constructed by means of a Gaussian kernel. We will call it *analysis pyramid*. Then, the corresponding pyramid for the synthetic texture - which we call *synthesis pyramid* - is built by *filling* its levels starting from the coarsest resolution, with values which are obtained by conditional sampling on the corresponding levels within the analysis pyramid. In order to enclose enough structural information in the conditioning procedure, a set of local descriptors is derived for each level of the analysis pyramid by applying a set of spatially selective filters. These could either correspond to steerable filters [10] or to a filter bank of oriented first and second Gaussian derivatives.

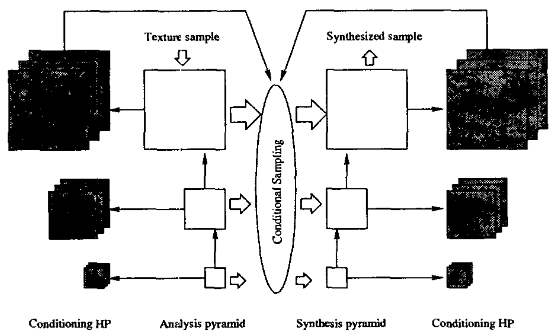


Fig. 1. MPTM: General scheme.

A *parent vector* is defined in each point:

$$\vec{V}(x,y) = [F_0^0(x,y), F_0^1(x,y), \dots, F_0^M(x,y)$$
$$F_1^0(\frac{x}{2},\frac{y}{2}), F_1^1(\frac{x}{2},\frac{y}{2}), \dots, F_1^M(\frac{x}{2},\frac{y}{2}) \dots$$
$$F_N^0(\frac{x}{2^N},\frac{y}{2^N}), F_N^1(\frac{x}{2^N},\frac{y}{2^N}), \dots, F_N^M(\frac{x}{2^N},\frac{y}{2^N})]$$

where $(N+1)$ is the number of pyramid levels, F_l^α is the *feature image* for orientation α and level l, and $(M+1)$ is the total number of orientations. A chain is defined across scales according to [6]:

$$\begin{aligned} p(\vec{V}(x,y)) &= p(\vec{V}_N(x,y)) \times p(\vec{V}_{N-1}(x,y)|\vec{V}_N(x,y)) \\ &\times p(\vec{V}_{N-2}(x,y)|\vec{V}_{N-1}(x,y),\vec{V}_N(x,y)) \\ &\times p(\vec{V}_0(x,y)|\vec{V}_1(x,y),\vec{V}_{N-1}(x,y),\dots\vec{V}_N(x,y)) \end{aligned} \quad (1)$$

where $\vec{V}_l(x,y)$ is the subset of the elements of $\vec{V}(x,y)$ concerning level l of the pyramid. The conditional distributions are estimated from observations. In [6], the non-parametric Parzen estimator [11] is used. Let $\vec{V}_l^k$ be the part of the parent vector collecting the features pertaining levels l to k. Then, the conditional probability of $\vec{V}_l(x,y)$ is estimated from the data as:

$$\begin{aligned} p(\vec{V}_l(x,y)|\vec{V}_{l+1}^N(x,y)) &= \frac{p(\vec{V}_l(x,y)|\vec{V}_l^N(x,y))}{p(\vec{V}_l(x,y)|\vec{V}_{l+1}^N(x,y))} \\ \approx & \frac{\sum_{x',y'} \varphi(\vec{V}_l^N(x,y),\vec{S}_l^N(x',y'))}{\sum_{x',y'} \varphi(\vec{V}_{l+1}^N(x,y),\vec{S}_{l+1}^N(x',y'))} \end{aligned} \quad (2)$$

where φ is the Parzen window function. The adopted definition for the window consists in an $(N+1) \times (M+1)$-dimensional box. The returned value is a certain constant z (different from zero) if the $\vec{S}$ vector falls inside the box pertaining to the $\vec{V}$ vector, and zero otherwise. Due to the definition of φ, sampling from $p(\vec{V}_l(x,y)|\vec{V}_{l+1}^N(x,y))$ can be done as follows. First, determine the set of all x', y' such that:

$$\varphi(\vec{V}_{l+1}^N(x,y),\vec{S}_{l+1}^N(x',y')) = z \quad (3)$$

Then, pick from them to set $\vec{V}_l(x,y) = \vec{S}_l(x',y')$. Equation (3) is used to build the *candidate set* $C_i(x,y)$ for each sample of the synthesis pyramid. In other words, the point in (x',y') of the analysis pyramid is a candidate for sampling for (x,y) in the synthesis pyramid if the parent vector of (x',y') falls inside the box pertaining to (x,y). Due to the shape of the Parzen window, this amounts to taking the distance between the corresponding elements of the parent vectors, and to compare it to a predefined *threshold* value, which represents the length of the box edge along the corresponding axis. More formally:

$$C_i(x,y) = \{(x',y') \ : \ D(\vec{V}_{l+1}^N(x,y),\vec{S}_{l+1}^N(x',y')) \le \vec{T}_i\} \quad (4)$$

$\vec{T}_i$ being the vector collecting the thresholds. The less-equal operator is intended to be applied to each vector element. The levels of the analysis pyramid are then sampled uniformly from among all regions having a parent structure which satisfies (4).

The generative procedure starts by filling the coarsest (low-pass) subband of the synthesis pyramid by copying the corresponding values within the analysis pyramid - or eventually tiling it in case the texture to be generated has a larger size. The sampling procedure is then successively applied to the lower pyramid levels.

3. DWT-BASED MPTM

The DWT-based approach is illustrated in fig. 2. The texture de-

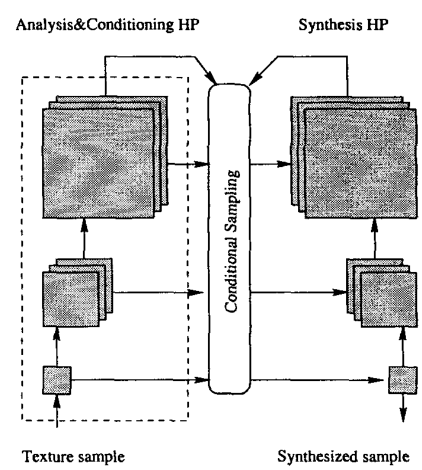


Fig. 2. DWT-based MPTP.

tails in each pyramid level are represented by the DWT wavelet coefficients, namely the analysis pyramid is replaced by DWT hyper-pyramid. The conditioning hyper-pyramid has been collapsed on the analysis one, which is thus used for *both* sampling and conditioning. The reason behind this is that the role of the feature images (conditioning HP) is to maintain the cross-correlation among subbands, representing different orientations, in order to preserve the structure of the texture. The DWT details are suitable for this purpose, even though probably not optimal. Figure 3 illustrates the construction of the analysis&conditioning HP. The 1D version is considered for simplicity of representation. In the block diagram,

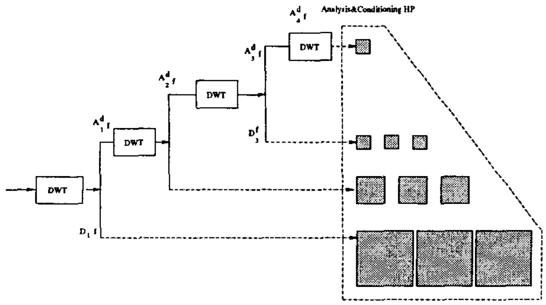


Fig. 3. DWT-based MPTP: analysis&conditioning HP.

DWT represents one level of transformation. The first DWT block results in the coarsest scale approximation $A_1^d f$ and details $D_1 f$. The DWT is iterated on the approximation $A_1^d f$, leading to the next finer scale subbands $A_2^d f$ and $D_2 f$, etc.

The synthesis algorithm articulates as follows. Let N be the height of the HP ($N = 4$ in the example), and let l be the level index. The signal approximation ($l = N$) and the first level of the HP ($l = N - 1$) are replicated in the synthesis HP. Levels from $l = N - 2$ are filled by conditional sampling. For each level l, the three detail subbands are sampled together in order to preserve inter-band dependencies. Depending on the texture structure, it may be convenient to stop the sampling procedure at a certain $l = \bar{l}$. In this case, the whole tree of descendants of each coefficient, which is the outcome of the sampling, is replicated in the synthesis HP. This corresponds to moving whole textural units while preserving high frequencies. The main drawback is that some *blocking* artifacts may appear, reflecting the pyramidal structure of the decomposition.

3.1. Conditional Sampling

The choice of the threshold vector $\vec{T}_l$ is crucial. Large values would introduce a high level of randomization in the feature space, which could compromise the preservation of the local structure. Conversely, too small values would *overconstrain* the system and result in just a local re-arrangement of the coefficients. The choice of threshold values is what determines the performances of the probability density estimator. As pointed out in [11], if the Parzen window φ is too large, then the estimated distribution would be a superposition of broad, slowly varying functions, corresponding to a very smooth, "out-of-focus" estimate of the true distribution. On the other end, highly peaked windows would generate a superposition of sharp pulses located at the samples, leading to an "erratic" noisy estimate of the distribution. In other words, in the first case the estimate would suffer from too little resolution, and in the second from too much statistical variability. If the power of non-parametric techniques is the ability to estimate *a priori* unknown distributions (with *enough samples*, convergence to any distribution is ensured), the main limitation is that the number of necessary samples may be very large. Furthermore, the demand for a large number of samples grows exponentially with the dimensionality of the feature space.

The parameter we have used as reference for the size of the box along the axis ν, corresponding to the feature image F_l^α, is the unbiased estimation for the root mean square value:

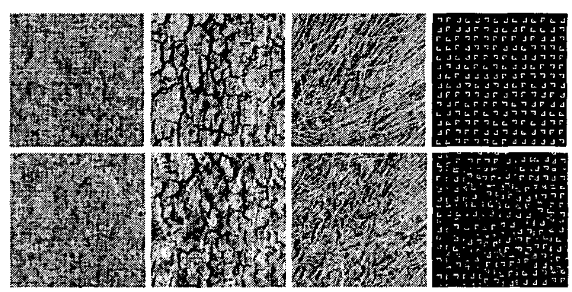
$$\sigma_l^\alpha = \sqrt{\frac{1}{N_l^\alpha - 1} \sum_{x,y} (F_l^\alpha(x,y) - \mu_l^\alpha)^2} \qquad (5)$$

μ_l^α and N_l^α being the expectation value and the number of samples in F_l^α, respectively. These values do not ensure the preservation of the structure for any texture. In some cases, results can be improved by an adequate weighting of the reference values. The Parzen window is one of the core parameters of our approach, and deserves further investigation.

4. RESULTS AND DISCUSSION

The test set consists of some textures taken from the Brodatz album [12]. Additional artificial textures have been included for sake of comparison with other methods [5, 13]. The limit on the number of decomposition levels is set by both the size of the input images and the length of the filters. For the examples shown it varies between $N = 3$ and $N = 4$ for original samples of 64×64 pixels each, and between $N = 4$ and $N = 5$ for the 128×128 ones. The integer version of the Daubechies' (9,7) filter [8] is used. The size of the Parzen window has been fixed according to eq. (5). The visual quality of the resulting textures could be further improved by a fine tuning of such a parameter.

Figures 4 and 5 illustrate performances. Fig. 4 (a) and (b) represent some natural textures. The first one is quite regular, and features mainly horizontal and vertical structures, making it well suited to the proposed approach. For these two samples, the synthesized images can be considered perceptually equivalent to the originals [1]. Fig. 4 (c) has pronounced directional features away from the horizontal/vertical directions. These are not always correctly preserved. Fig. 4 (d) shows a synthesis result on a classical



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

Fig. 4. Synthesis results. First line: original images. Second line: synthesized textures.

[1]No standardized psycho-physical evaluation has been performed.

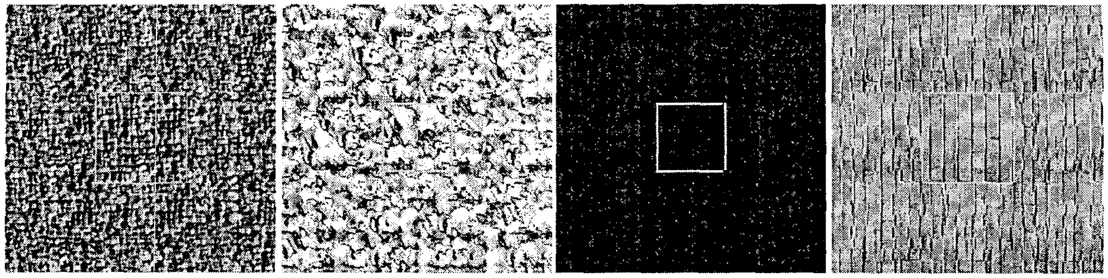


Fig. 5. Texture extrapolation. The original texture is pasted at the center.

texton image. The textons are placed at independent random non-overlapping positions within the image. Some artifacts are present in the form of artificially generated gray levels that were absent in the original bi-level image. This is due to the filtering process, which expands the original number of gray levels. As previously mentioned, the proposed algorithm is very easily extensible to texture extrapolation. In such case, each level of the synthesis HP has a bigger size than the corresponding one in the analysis HP. In order to fill the last levels ($l = N, N-1$) of the synthesis HP, the corresponding levels of the analysis HP are tiled. Figure 5 shows some examples of texture *extrapolation*. The original sample is pasted at the center, and the rest of the image is filled with the synthetic texture. To our opinion, image 5(d) is particularly representative of the MPTM philosophy: the synthetic image looks like a wall built by a different arrangement of the bricks. Performances can be improved by fine tuning of the parameters of both the DWT (filter, depth) and the sampling. Among these, the shape and size of the Parzen window and the level $\bar{l}$, at which it may be convenient to replace the sampling with the replication of the descendants, are particularly important because they control the degree of randomization introduced by the algorithm.

5. CONCLUSIONS

We propose a multi-resolution probabilistic texture modeling technique entirely based on the integer DWT. The input image is first analyzed by measuring the joint occurrence of texture features at multiple resolutions. Then, a new texture is generated by sampling successive frequency bands from the input texture, conditioned on the joint occurrence of similar features at lower spatial frequencies. The core of the algorithm is the sampling procedure, controlling the degree of randomization introduced in the feature space. Results on stochastic textures are comparable with other state-of-the-art techniques. For structured textures, performance depends on the orientation features and structure size. The low computational complexity and the possibility of implementing the transform with integer arithmetic make the proposed technique highly appealing for many applications. We aim at integrating the DWT-based algorithm within a model-based coding system.

6. REFERENCES

[1] R.L. De Valois, R.L. Albrecht, and L.G. Thorrell, "Spatial frequency selectivity in macaque visual cortex," *Viosion Research*, vol. 22, pp. 545–559, 1982.

[2] G.C. De Angelis, I. Ohzawa, and R.D. Freeman, "Spatiotemporal organization of simple cell receptive field in the cat's visual cortex. i. general characteristics and postnatal development," *J. Neorophysiol.*, vol. 69, pp. 1091–1117, 1993.

[3] J. Portilla and E.P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int'l Journal of Computer Vision*, vol. 40(1), pp. 49–71, Dec. 2000.

[4] S. Zhu, Y. Wu, and M. Mumford, "Filters, random fields and maximum entropy (frame) - towards the unified theory for texture modeling," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 1996.

[5] J. S. De Bonet, "Multiresolution sampling procedure for analysis and synthesis of texture image," in *Computer Graphics*. ACM SIGGRAPH, 1997, pp. 361–368.

[6] J. S. De Bonet and P. Viola, "A non-parametric multi-scale statistical model for natural images," *Advances in Neural Information Processing*, vol. 10, 1997.

[7] G. Menegaz, *Model-based coding of multi-dimensional data with applications to medical imaging*, Ph.D. thesis, Signal Processing Laboratory (LTS), Swiss Federal Institute of Technology (EPFL), May 2000.

[8] A.R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *Appl. Comput. Harmon. Anal.*, vol. 5, no. 3, pp. 332–369, 1998.

[9] Julien Reichel, *Complexity related aspects of image compression*, Ph.D. thesis, Swiss Federal Institute of Technology (EPFL), Feb. 2001.

[10] E.P. Simoncelli, W.T. Freeman, and E.H. Adelson, "Shiftable multi-scale transforms," *IEEE Trans. on Information Theory*, vol. 38, pp. 587–607, 1992.

[11] R.O. Duda and P.E. Hart, *Pattern classification and scene analysis*, John Wiley and sons, 1973.

[12] P. Brodatz, *Textures: a photgraphic album for artits and designers*, Dover, New York, 1966.

[13] J. Portilla and E. Simoncelli, "Texture modeling and synthesis using joint statistics of complex wavelet coefficients," in *Proc. of the IEEE Workshop on Statistical and Computational Theories of Vision*, June 1999.