DISCRETE WAVELET TRANSFORM WITH OPTIMAL JOINT LOCALIZATION FOR DETERMINING THE NUMBER OF IMAGE TEXTURE SEGMENTS

P. Tay, J.P. Havlicek, and V. DeBrunner

School of Electrical and Computer Engineering
University of Oklahoma, Norman, OK USA
{ptay,joebob,vdebrunn}@ou.edu

ABSTRACT

Accurate estimation of the number of textured regions that are present in an image is one of the most difficult aspects of the unsupervised texture segmentation problem. In this paper we introduce a new approach for estimating the number of regions in an image without *a priori* information. Using a novel discrete-discrete uncertainty measure defined on equivalence classes of signals, we design a localized separable 2-D wavelet transform. By clustering in a feature space defined by the wavelet coefficients computed over disjoint blocks in the image, we obtain high quality estimates for the number of textured regions present in an image. Compared to a previously reported algorithm based on the eight-point Daubechies wavelet, this new approach tends to produce clusters with improved between-cluster separations.

1. INTRODUCTION

Numerous texture segmentation techniques have been proposed in the literature, including, *e.g.*, approaches based on wavelet analysis [1, 2], filter banks [3, 4], deterministic annealing [5], and stochastic models [6–8]. Determining the number of texture regions present in an image without *a priori* information is one of the most challenging problems that must be addressed in developing an unsupervised texture segmentation algorithm. We recently described a technique where the number of regions was determined by partitioning the image into small disjoint blocks and computing separable 2-D discrete wavelet transforms on each block [9], where the well-known Daubechies D_8 wavelet was used [10]. Feature vectors were constructed from the wavelet transform coefficients and statistical clustering was performed in the resulting feature space to build a full dendrogram. Finally, a validation criterion was applied to select the most likely clustering configuration and estimate the number regions in the image. This technique can be combined with a partially supervised approach to estimate the number of regions and realize a fully unsupervised algorithm.

In this paper we describe a novel uncertainty measure which, for a signal that is discrete and finitely supported in both time (or space) and frequency, quantifies the signal's joint localization in the two domains. This new measure is invariant under translations and modulations. It also bears analogies to the well-known continuous Heisenberg-Weyl uncertainty principle and admits an intuitively satisfying interpretation in terms of statistical variance. By introducing several reasonable constraints, we obtain a 1-D lowpass analysis filter with optimal joint time-frequency localization and use it to construct quadrature mirror filters for performing a separable discrete 2-D wavelet transform. Compared to the previous work reported in [9], the new wavelet transform provides improved localization manifest as increased between-cluster distances, which enhance the robustness of our overall strategy for estimating the number of regions that are present. The technique is demonstrated on juxtapositions of Brodatz-like textures.

2. A JOINTLY LOCALIZED DISCRETE WAVELET

Consider a continuous-time signal $x : \mathbb{R} \longrightarrow \mathbb{C}$, $x \in C^1$, with $\lim_{t\to\pm\infty} x(t) = 0$, and also its Fourier transform $X(f) = \int_{\mathbf{R}} x(t)e^{-j2\pi ft}\,dt$. The simultaneous or *joint* time-frequency localization of $x(t)$ is bounded below by the well-know Heisenberg-Weyl uncertainty principle (HUP) $\Delta_t\Delta_f \geq \frac{1}{2}$, where

$$\Delta_t = \frac{1}{||x||_{L^2}^2}\int_{\mathbf{R}}(t-\langle t\rangle)^2|x(t)|^2\,dt, \quad (1)$$

$$\Delta_f = \frac{1}{||X||_{L^2}^2}\int_{\mathbf{R}}(f-\langle f\rangle)^2|X(f)|^2\,df \quad (2)$$

and where the mean time $\langle t\rangle$ in (1) is defined by $\int_{\mathbf{R}} t|x(t)|^2\,dt/||x||_{L^2}^2$. The mean frequency $\langle f\rangle$ in (2) is defined similarly.

We consider the unit L^2-norm functions $|x(t)|^2/||x||_{L^2}^2$ and $|X(f)|^2/||X||_{L^2}^2$ as densities characterizing, respectively, the distributions of signal energy in time and in frequency. Δ_t and Δ_f then admit intuitively appealing formal interpretations as statistical variances quantifying how well $x(t)$ is localized in each domain. Additional appealing properties of the HUP include the facts that

- it is invariant under the dual operations of translation in time (complex modulation in frequency) and translation in frequency (complex modulation in time); intuitively, we do not expect that simple translation should have an effect on localization, and
- as is well known, the HUP lower bound on joint localization is realized uniquely by a nontrivial and useful family of functions, *viz.*, the Gabor elementary functions.

Many modern digital image processing applications are concerned exclusively with signals that are discrete and finitely supported, however. It is an unfortunate fact that the HUP cannot be applied to such signals. Moreover, the frequency representation of greatest practical interest is that delivered by the discrete Fourier transform (DFT). Thus practical applications depend on signal representations that are discrete and finitely supported both in time (or space) and in frequency. There is a critical need to develop corresponding *discrete-discrete* uncertainty measures capable of quantifying joint localization for such representations.

2.1. Discrete-Discrete Uncertainty Measures

For a finitely supported discrete signal, Donoho and Stark considered discrete-discrete uncertainty measures based on the counting measure of the signal's time and frequency support [11]. DeBrunner, Özaydin, and Przebinda have also developed entropy based discrete-discrete measures related to the Hirschmann uncertainty principle [12]. In addition to being novel and useful, all of these measures are invariant under translations and modulations. However, with relation to the HUP, none of them admit intuitive interpretation in terms of statistical variance. Moreover, these measures are invariant under *arbitrary* permutations of the signal samples, whereas the HUP is not. In this section we briefly describe a new discrete-discrete uncertainty measure that is invariant under translation and modulation and bears strong analogy to the HUP.

Let $x[n]$ map $[0, N-1] \longrightarrow \mathbb{C}$ and let $X[k]$ be the N-point DFT of $x[n]$. Naïve discretization of (1) and (2) yields

$$\Delta_n = \frac{1}{||x||_{\ell^2}^2} \sum_{n=0}^{N-1} (n-\mu)^2 |x[n]|^2, \tag{3}$$

$$\Delta_\omega = \frac{1}{||X||_{\ell^2}^2} \sum_{k=0}^{N-1} (k-\nu)^2 |X[k]|^2, \tag{4}$$

where $\mu = (\sum_{n=0}^{N-1} n|x[n]|^2)/||x||_{\ell^2}^2$ and $\nu = (\sum_{k=0}^{N-1} k|X[k]|^2)/||X||_{\ell^2}^2$.

While the localization measures (3) and (4) may be interpreted in terms of variance and are therefore appealing, their chief shortcoming is that they are neither translation invariant nor modulation invariant. The key to overcoming this difficulty is to define related measures not on discrete signals, but rather on equivalence classes of discrete signals. We define an equivalence relation $x \sim y$ between discrete N-point signals $x[n]$ and $y[n]$ if $\exists\, p, q, r \in \mathbb{Z}$ such that

$$y[n] = e^{j\frac{2\pi}{N}(qn+r)} x[(n-p) \bmod N]. \tag{5}$$

This leads to the important result that $x[n] \sim y[n]$ if and only if $X[k] \sim Y[k]$; we define the equivalence class of $x[n]$ by $[x] = \{y[n] : y \sim x\}$.

Note that the equivalence class $[x]$ contains exactly those signals $y[n]$ that are obtained from $x[n]$ by applying translations, modulations, and constant phase shifts – none of which should be expected to have an effect on signal localization. Also, the members of $[x]$ are precisely the inverse DFT's of the elements of $[X]$. For any $y \in [x]$, we define time (space) and frequency uncertainty measures according to

$$\sigma_{n,[x]}^2 = \min_{[x]} \Delta_n, \tag{6}$$

$$\sigma_{\omega,[x]}^2 = \min_{[X]} \Delta_\omega, \tag{7}$$

where Δ_n and Δ_ω were given in (3) and (4). These measures admit an interpretation in terms of variance, are translation and modulation invariant, and are generally affected by arbitrary permutations of the signal samples. The joint localization of any $y \in [x] = [y]$ is then quantified by the discrete-discrete measure $\gamma_{N,y}^2 = \sigma_{n,[y]}^2 \sigma_{\omega,[y]}^2$.

For finitley supported discrete signals taking values with finite magnitudes, one can easily establish that $\gamma_{N,y}^2 \geq 0$. Several signals realize the equality in this lower bound, including the Kronecker delta, the constant sequence, and pure sinusoids. Therefore, in using $\gamma_{N,y}^2$ to obtain meaningful optimally localized signals, it is generally necessary to restrict the class of signals under consideration and exclude the signals for which $\gamma_{N,y}^2 = 0$ by imposing additional constraints. With regards to the HUP, it should be noted that the analogous special cases (Dirac delta, *etc.*) were ruled out by the requirements in (1) and (2) that $x(t)$ be continuous ($\in C^1$) and that $\lim_{t\to\pm\infty} x(t) = 0$.

2.2. Optimal Wavelet QMF

In this section we apply the discrete-discrete uncertainty measure $\gamma_{N,h}^2$ on a restricted class of signals to design an optimally localized separable 2-D wavelet quadrature mirror filter bank (QMF). The fundamental problem is to design the impulse response $h[n]$ of the corresponding 1-D low-pass analysis filter to minimize $\gamma_{N,h}^2$ subject to the following constraints:

C1. N even

C2. $h[n]$ real valued

C3. $||h||_{\ell^2} = 1$

C4. $|H[0]| = \sqrt{2}$

C5. $|H[\frac{N}{2}]| = 0$

C6. $|H[k]| \leq \sqrt{2}$

C7. $\varphi[k] = \begin{cases} -\frac{\pi}{N}k & \text{for } k = 0, 1, \ldots, \frac{N}{2}, \\ \frac{\pi}{N}(N-k) & \text{for } k = \frac{N}{2}+1, \ldots, N-1, \end{cases}$

where $H[k] = |H[k]|e^{-j\varphi[k]} = \sum_{n=0}^{N-1} h[n] e^{-j\frac{2\pi}{N}nk}$.

Constraints (C1)-(C6) are well-known in the wavelet literature [10] and imply that

$$|H[\tfrac{N}{2}-1]| = \sqrt{\tfrac{N-2}{2} - \sum_{k=1}^{\frac{N-4}{2}} |H[k]|^2}, \tag{8}$$

while constraints (C1) and (C2) imply that

$$|H[k]| = |H[N-k]| \text{ for } k = 1, 2, \ldots, \tfrac{N}{2}-1. \tag{9}$$

The spectral phase constraint (C7) was adopted for several reasons. First, it corresponds to the spectral phase of the Haar scaling function, which is generally well localized in time (space) and is the only admissible low-pass analysis filter of length $N = 2$ under constraints (C1)-(C6). Second, a possible minimizer for the DTFT-based joint uncertainty measure (a *discrete-continuous* measure) given by Calvez and Vilbé in [13] had linear phase. The form given in constraint (C7) is then justified by observing that the DFT corresponds to uniformly spaced samples of the DTFT. Finally, linear phase in the low-pass analysis filter assures that the resulting wavelet QMF's can be cascaded to achieve different resolutions without the need for phase compensation. One unavoidable consequence of condition (C7) is that the resulting QMF cannot be orthogonal (unless $h[n]$ is the Haar scaling function); we consider that a low-pass analysis filter $h[n]$ satisfying constraint (C7) sacrifices orthogonality in the interest of obtaining improved joint localization.

A numerical optimization was implemented to minimize $\gamma_{N,h}^2$ subject to constraints (C1)-(C7) for filter lengths N = 2, 4, 6, and 8. The impulse response of the resulting length $N = 8$ low-pass QMF is given by

$$\begin{aligned} h[0] &= -0.0131588203123 &&= h[7] \\ h[1] &= -0.0284383229781 &&= h[6] \\ h[2] &= 0.04364111877368 &&= h[5] \\ h[3] &= 0.70506280570341 &&= h[4]. \end{aligned}$$

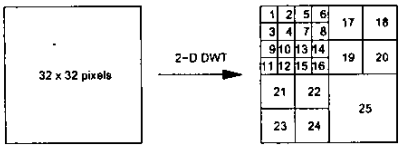


Fig. 1. The 25 channels of the 2-D, three-level wavelet decomposition.

For comparison, the joint uncertainties $\gamma^2_{8,h}$ of the optimal filter, the 8-point Haar low-pass filter, and the 8-point Daubechies low-pass filter are 0.4843, 0.5214, and 0.9923, respectively.

The 1-D QMF was completed by designing the impulse response of the high-pass analysis filter according to $g[n] = (-1)^n h[n]$. We implemented a 2-D separable discrete wavelet transform by sequentially performing 1-D convolution along image rows with the appropriate filter, down sampling by a factor of two, and subsequently performing 1-D convolution along the resulting columns and again down sampling by a factor of two. Edge effects were handled by symmetric reflection.

3. ESTIMATING THE NUMBER OF IMAGE REGIONS

We begin by partitioning the image into equal-sized disjoint blocks. The block size should be chosen small enough to ensure that typical segments are covered by several blocks and also to limit the fraction of blocks that span multiple segments. Concomitantly, the blocks must be large enough to yield meaningful wavelet transform coefficients. In the interest of clarity we henceforth assume 256×256 grayscale images partitioned into blocks B_i of 32×32 pixels each, where $1 \leq i \leq M$ and $M = 64$. We index the blocks in row major order so that B_1 and B_2 are the first and second leftmost blocks on the first block row, whereas B_{64} is the rightmost block on the last block row of the image.

The jointly localized 2-D wavelet transform described in Section 2 is applied independently on each block to obtain a three-scale multiresolution analysis describing the block's texture content. Since it is well established that low frequencies dominate virtually all real images, we construct the multiresolution analysis to be denser in the lower spatial frequencies. This produces a 25 channel subband decomposition of each block as depicted in Fig. 1.

3.1. Feature Space

Since multiple disjoint segments sharing a single texture pattern may generally be present in the image, we combine spatial information with the wavelet transform coefficients to construct a feature vector for each image block B_i. For each $i \in [1, M]$ and each $k \in [1, 25]$, let $e_{i,k}$ denote the squared ℓ_2-norm of the wavelet coefficients in the k^{th} subband of image block B_i. We describe block B_i by constructing a wavelet domain feature vector $\mathbf{e}_i$ according to

$$\mathbf{e}_i = [e_{i,1}\ e_{i,2}\ \ldots\ e_{i,25}]^T.$$

Let $\mathcal{F} = \{\mathbf{e}_i : i \in [1, M]\}$. Thus, $\mathcal{F}$ is a 25-D feature space that contains M vectors, each describing one block from the original image.

Spatial information is incorporated by augmenting the feature vectors with two additional coordinates. Let r_i and c_i denote, respectively, the average row coordinate and average column coordinate for pixels in block B_i. Let $\mathcal{C} = \{[r_i\ c_i]^T : i \in [1, M]\}$. Then $\mathcal{C}$ contains vectors that describe the spatial centroids of the M image blocks B_i. The feature space is given by $\mathcal{F} \times \mathcal{C}$. In this feature space, image block B_i is described by the vector $\mathbf{w}_i = [\mathbf{e}_i^T\ r_i\ c_i]^T$.

For each $k \in [1, 27]$, the collection of the k^{th} entries from all M vectors $\mathbf{w}_i \in \mathcal{F} \times \mathcal{C}$ defines the k^{th} *feature*. To minimize the possibility that one or a few features with relatively large numerical values might dominate the segmentation procedure, we normalize each feature independently. For feature k, the normalization consists of first computing the sample standard deviation of the feature and then dividing the k^{th} entry of each vector $\mathbf{w}_i$ by this value. We use the notation $\mathcal{F}' \times \mathcal{C}'$ to denote the normalized feature space.

3.2. Nearest Neighbor Clustering

The nearest neighbor clustering algorithm (NNC) is used to partition the normalized feature space $\mathcal{F}' \times \mathcal{C}'$. Initially, each of the M feature vectors is considered to be a cluster. The algorithm iterates through $M - 1$ passes. At each pass, we merge the two clusters that are closest with respect to the distance metric described in (11) below. Thus, there are M clusters prior to the first pass and only one cluster remains after pass $M - 1$. Let

$$\delta(\mathbf{w}_i, \mathbf{w}_j) = \lambda d(\mathbf{e}_i, \mathbf{e}_j) + (1 - \lambda) d([r_i, c_i]^T, [r_j, c_j]^T), \quad (10)$$

where $0 \leq \lambda \leq 1$ and $d(\cdot, \cdot)$ is the usual Euclidean metric. The term λ appearing in (10) weights the relative contributions of the wavelet coefficient energies in $\mathcal{F}'$ and the spatial position information in $\mathcal{C}'$. We typically take $\lambda = 0.8$.

In a given pass of the NNC algorithm, let L_j denote the number of feature vectors contained in cluster C_j and induce an arbitrary ordering on these feature vectors so that

$$C_j = \{\mathbf{w}_{j,1}\ \mathbf{w}_{j,2}\ \ldots\ \mathbf{w}_{j,L_j}\}.$$

We define the distance between clusters C_j and C_k by

$$\Delta(C_j, C_k) = \min_{\substack{p \in [1, L_j] \\ q \in [1, L_k]}} \delta(\mathbf{w}_{j,p} \mathbf{w}_{k,q}). \quad (11)$$

$\Delta(C_j, C_k)$ defines the *closeness* of clusters C_j and C_k by the distance between their nearest elements with respect to the metric δ.

When it terminates after $M - 1$ iterations, the NNC algorithm delivers M cluster configurations $\Gamma_M \ldots \Gamma_1$, where k clusters are present in configuration Γ_k. We choose one of these as the final clustering result by applying a validation criterion to quantify the "goodness" of each configuration. Typically, for some K considered to be the maximum number of segments that might be present in the image, validation is applied only to configurations Γ_k for $k \in [1, K]$. For the examples in Section 4, we took $K = 10$.

The validation criterion is applied to configuration Γ_k as follows. Using the distance metric (10), we first compute the centroid of each of the clusters in the configuration. Then we compute $\overline{C}_k$, the average distance between any two distinct centroids. The average within-cluster distance $\overline{W}_k$ is the average distance from all the feature vectors in $\mathcal{F}' \times \mathcal{C}'$ to the centroid of their respective clusters. The goodness of Γ_k is then defined by the ratio $R_k = \overline{C}_k / \overline{W}_k$. Ideally, we would like for the average between cluster distance to be large and the average within-cluster distance

Wavelet	Distance between feature vectors from typical blocks in different regions			
Optimal	70.2	98.9	135.9	50.1
D_8	52.2	93.7	93.9	29.6

Table 1. Typical squared distances between feature vectors corresponding to blocks from different textures computed using localized wavelet and D_8.

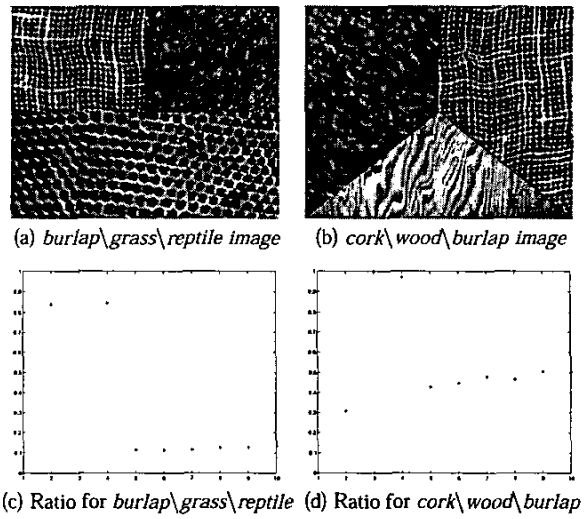
to be small. Hence our validation criterion selects as the final result that configuration Γ_k that maximizes R_k. For our estimate $\mathcal{N}$ of the number of texture segments that are present in the image, we use the number of clusters in the final clustering result, which is given by $\mathcal{N} = \arg\max_{k \in [1,K]} R_k$.

4. EXAMPLES & CONCLUSIONS

Two three-texture images are given in Fig. 2(a) and (b). The ratio R_k computed for these images is graphed as a function of k in Fig. 2(c) and (d), where it can be seen that the proposed technique correctly selects the value $\mathcal{N} = 3$ for the number of texture segments that are present in each case.

Consider the image of Fig. 2(a). For typical blocks B_i and B_j located in different textured regions, we calculated the square of the Euclidean distance between feature vectors $\mathbf{e}_i$ and $\mathbf{e}_j$ computed using both the localized wavelet described in Section 2 and the eight-point Daubechies wavelet as in [9]. Several values of this squared distance are shown in Table 1. As we asserted in Section 1, these results indicate that the new wavelet proposed in this paper generally results in better separation between feature vectors corresponding to blocks from differing textures, which leads to increased between-cluster distances that enhance the robustness of the estimates $\mathcal{N}$ delivered by the algorithm. These results are typical of those that we have obtained over a large variety of two- and three-texture images.

This paper has made several original contributions. First, we introduced a new uncertainty measure quantifying the joint localization of a finitely supported signal in discrete time and discrete frequency. This measure is attractive because it bears strong analogies to the continuous Heisenberg-Weyl uncertainty principle. Realizing the need to define the measure on equivalence classes of signals was the key to achieving both shift invariance and modulation invariance. A new optimally localized wavelet transforem was designed and we demonstrated its usefulness in estimating the number of textured regions present in an image. This is important in that it provides a means of deriving fully unsupervised texture segmentation algorithms from a variety of partially supervised approaches that are currently available.



(a) *burlap\grass\reptile image* (b) *cork\wood\burlap image*

(c) Ratio for *burlap\grass\reptile* (d) Ratio for *cork\wood\burlap*

Fig. 2. Three-texture examples.

5. REFERENCES

[1] R. Porter and N. Canagarajah, "A robust automatic clustering scheme for image segmentation using wavelets," *IEEE Trans. Image Proc.*, vol. 5, no. 4, pp. 662–665, April 1996.

[2] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Proc.*, vol. 4, no. 11, pp. 1549–1560, Nov. 1995.

[3] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 1, pp. 55–73, Jan. 1990.

[4] D. Dunn and W.E. Higgins, "Optimal gabor filters for texture segmentation," *IEEE Trans. Image Proc.*, vol. 4, no. 7, pp. 947–964, July 1995.

[5] T. Hofmann, J. Puzicha, and J.M. Buhmann, "Unsupervised texture segmentation in a deterministic annealing framework," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 8, pp. 803–818, Aug. 1998.

[6] D. A. Langan, J. W. Modestino, and J. Zhang, "Cluster validation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Proc.*, vol. 7, no. 2, pp. 180–195, Feb. 1998.

[7] J. Chen and A. Kundu, "Unsupervised texture segmentation using multichannel decomposition and hidden markov models," *IEEE Trans. Image Proc.*, vol. 4, no. 5, pp. 603–619, May 1995.

[8] C. Kervrann and F. Heitz, "A markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics," *IEEE Trans. Image Proc.*, vol. 4, no. 6, pp. 856–862, June 1995.

[9] J. P. Havlicek and P.C. Tay, "Determination of the number of texture segments using wavelets," in *Proc. 16th Conf. Appl. Math.*, Edmond, OK, February 23-24 2001, pp. 61–70.

[10] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 51, pp. 909–996, 1988.

[11] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, June 1989.

[12] V. DeBrunner, M. Özaydin, and T. Przebinda, "Resolution in time-frequency," *IEEE Trans. Signal Proc.*, vol. 47, no. 3, pp. 783–788, March 1999.

[13] L. Calvez and P. Vilbé, "On the uncertainty principle in discrete signals," *IEEE Trans. Circuits, Syst.-II: Analog, Digital Sig. Proc.*, vol. 39, no. 6, pp. 394–395, June 1992.