The Multicomponent AM–FM Image Representation

Joseph P. Havlicek, David S. Harding, and Alan C. Bovik

***Abstract*—We compute AM–FM representations for multicomponent, nonstationary images using a statistical component model. Components are isolated with a filterbank comprising frequency and orientation selective channels. The modulating functions for each component are estimated from the channel responses using localized nonlinear operators followed by optimal MMSE estimators. We also demonstrate reconstruction from the representation.**

I. Introduction

The efficacy of AM–FM models for instantaneous frequency estimation and nonstationary signal and image characterization has recently been intensely studied [1]–[5]. A multidimensional version of the Teager–Kaiser operator has been used to compute AM–FM models of single-component images [1], [5], and related algorithms have been applied to more general images [2], [4], [6]. *Multicomponent* AM–FM modeling techniques represent images as sums of locally coherent complex-valued components, each of the form

$$t(\mathbf{x}) = a(\mathbf{x}) \exp[j\varphi(\mathbf{x})] \tag{1}$$

where $(\mathbf{x}) = [x_1, x_2]^T$, t: $\mathbb{R}^2 \to \mathbb{C}$, a: $\mathbb{R}^2 \to [0, \infty)$, and φ: $\mathbb{R}^2 \to \mathbb{R}$, or as sums of locally coherent real-valued components $s(\mathbf{x}) = a(\mathbf{x})\cos[\varphi(\mathbf{x})]$. We refer to $a(\mathbf{x})$ and $\nabla\varphi(\mathbf{x})$ as a component's amplitude and frequency modulating functions. By *locally coherent*, we mean that the modulating functions vary smoothly in the sense of having bounded first-order Sobolev norms (some advantages of AM–FM modeling may not be realized with components that are fractal, self-similar, or extraordinarily discontinuous in nature; such are termed *incoherent* in this sense). Local coherency may be quantified on a pointwise basis by the instantaneous bandwidth [7] as follows:

$$B(\mathbf{x}) = \left\{ \mathrm{Im}\left[\frac{\nabla t(\mathbf{x})}{j t(\mathbf{x})} \right] \right\}^2 = \frac{|\nabla a(\mathbf{x})|^2}{|a(\mathbf{x})|^2}. \tag{2}$$

If a multicomponent decomposition exists in a region such that $B(\mathbf{x})$ is appreciably smaller for the individual components than for the image, then we call the image multicomponent in that region. Consider the image of Fig. 1(a). One decomposition into two components is shown in the (b) and (c) parts of the figure. Fig. 1(d) histograms $\sqrt{B(\mathbf{x})}$ for the composite image, whereas $\sqrt{B(\mathbf{x})}$ is histogrammed for the individual components in Fig. 1(e) and (f). The image is clearly multicomponent since the two-component interpretation reduces $\sqrt{B(\mathbf{x})}$ by approximately two orders of magnitude.

The purpose of this correspondence is to present late-breaking research on the multicomponent demodulation problem by proposing a practical technique for computing the multicomponent AM–FM representation

$$\{\tilde{a}_k(m,n), \nabla\tilde{\varphi}_k(m,n)\}_{k\in[1,K],(m,n)\in[0,M-1]\times[0,N-1]} \tag{3}$$

of an M-column by N-row K-component discrete-domain image and demonstrate image reconstruction from the representation. Although computation of the AM–FM representation for complicated, natural multipartite images is a difficult problem, we present dramatic preliminary results, where the essential structure, features, and information content of such an image are clearly captured using only a small number of AM–FM components.

Note that $t(\mathbf{x})$ in 1 and the corresponding real component $s(\mathbf{x})$ are uniquely related through $s(\mathbf{x}) = \mathrm{Re}[t(\mathbf{x})]$ and $t(\mathbf{x}) = s(\mathbf{x}) + j\mathcal{H}[s(\mathbf{x})]$, where $\mathcal{H}[\cdot]$ indicates the 2-D Hilbert transform acting in the $\mathbf{e}_x = [1,0]^T$ direction, defined by

$$\mathcal{H}[s(\mathbf{x})] = \frac{1}{\pi}\int_{\mathbb{R}} s(\mathbf{x} - y\mathbf{e}_x)\frac{dy}{y} = s(\mathbf{x}) * \frac{\delta(\mathbf{x}^T\mathbf{e}_y)}{\pi\mathbf{x}^T\mathbf{e}_x}. \tag{4}$$

The integral is interpreted as a Cauchy principle value, $\mathbf{e}_y = [0,1]^T$, and $\delta(\cdot)$ is the Dirac delta. The Fourier transforms of $s(\mathbf{x})$ and $\mathcal{H}[s(\mathbf{x})]$ are related by $\mathcal{F}\{\mathcal{H}[s(\mathbf{x})]\} = -j\mathrm{sgn}[\Omega^T\mathbf{e}_x]S(\Omega)$, implying that the spectrum of the complex image $t(\mathbf{x})$ is supported only in quadrants I and IV of the $\Omega = [u,v]^T$ frequency plane. We refer to $t(\mathbf{x})$ as the *analytic image* associated with $s(\mathbf{x})$ [6]. In $\nabla\varphi(\mathbf{x})$, we have an unambiguous definition for the *instantaneous frequencies* of $t(\mathbf{x})$, and we define these to be the instantaneous frequencies of $s(\mathbf{x})$.

A single component of the form (1) can be demodulated by the local nonlinear algorithm

$$a(\mathbf{x}) = |t(\mathbf{x})|, \tag{5}$$

$$\nabla\varphi(\mathbf{x}) = \mathrm{Re}\left[\frac{\nabla t(\mathbf{x})}{j t(\mathbf{x})}\right] \tag{6}$$

which is *exact* for *any* general complex-valued component [2], [4]. Equations (6) and (2) should be compared. While *any* image can be represented exactly with a single component, there is no guarantee that such representations admit smooth modulating functions. Indeed, images exhibiting a multipartite character, which are of great practical interest, are better modeled as the real part of

$$t(\mathbf{x}) = \sum_{i=1}^{K} a_i(\mathbf{x})\exp[j\varphi_i(\mathbf{x})] \tag{7}$$

in which case, the nonlinear algorithm (5) and (6) suffers from cross-component interference. Hence, it is necessary to separate the individual components from one another. We accomplish this with a multiband filterbank, which must be sufficiently spectrally localized to prevent interference between components but also spatially localized to effectively capture nonstationarities.

The design of such a filterbank using a wavelet-like tesselation of Gabor functions, which optimally realize the uncertainty principle lower bound on conjoint spatio-spectral localization, appears elsewhere [2], [6], [8]. The isotropic, scaled, translated, unity L^2-norm channel filter is

$$g_m(\mathbf{x}) = \frac{1}{\sigma_m\sqrt{2\pi}}\exp\left[-\frac{1}{4\sigma_m^2}\mathbf{x}^T\mathbf{x}\right]\exp\left[j2\pi\Omega_m^T\mathbf{x}\right] \tag{8}$$

with radial center frequency $r_m = |\Omega_m|$ and orientation $\theta_m = \arg[\Omega_m]$. The η-peak radial octave bandwidth is $B = \log_2[\{r_m + \sqrt{-\ln\eta}/(2\pi\sigma_m)\}/\{r_m - \sqrt{-\ln\eta}/(2\pi\sigma_m)\}]$, and the η-peak orientation bandwidth is $\Theta = 2\arctan\sqrt{\gamma}$, where $\gamma = (2^B-1)^2/(2^B+1)^2$. We arrange the filters on a polar grid such that the η-peak contours of any four adjacent filters intersect at a single frequency. Radial center frequencies along rays progress geometrically with ratio R, beginning at r_0, and the angular spacing between rays is $\Lambda = 2\arcsin[(4R)^{-\frac{1}{2}}\{(R^2+1)(\gamma-1)+2R(\gamma+1)\}^{\frac{1}{2}}]$. For the examples in this correspondence, $r_0 = 9.6$ cycles per image, $R = 1.8$, $B = 1$

Manuscript received November 21, 1994; revised November 16, 1995. This work was supported in part by a grant from the Texas Advanced Research Projects Agency and by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant F49620-93-1-0307.

The authors are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA.

Publisher Item Identifier S 1057-7149(96)04182-6.

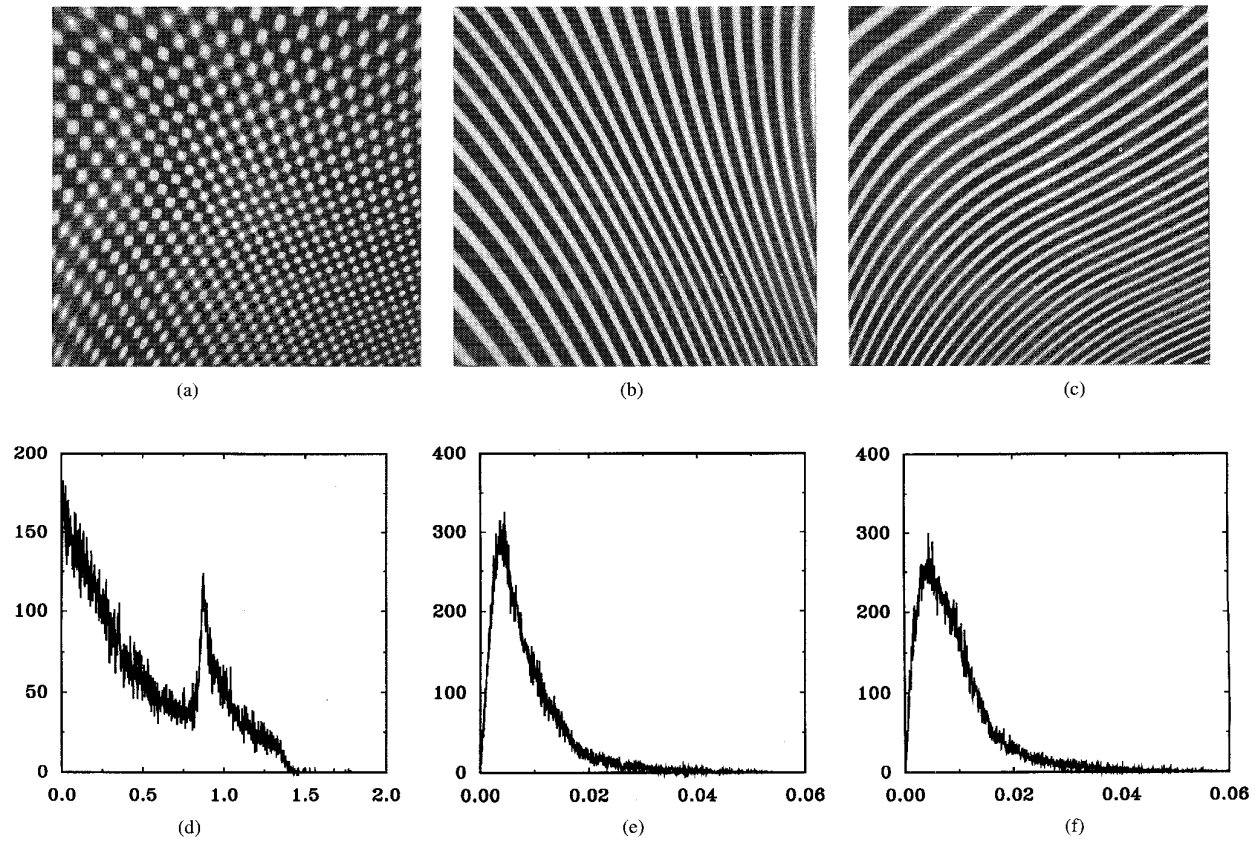


Fig. 1. Nonstationary multicomponent synthetic image showing one possible decomposition into highly locally coherent components. (a) Composite image. (b) Component one. (c) Component two. (d) Histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for the composite image; (e) histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for component one; (f) histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for component two. Note the difference in the abscissa scale between the (d) part of the figure and the (e) and (f) parts, which clearly indicates the multipartite nature of this image.

octave, and $\eta = \frac{1}{2}$, implying $\gamma = \frac{1}{9}$ and $\Theta \approx 38.9424^\circ$. Fig. 2 depicts this filterbank in the frequency domain.

Provided the filterbank design is such that at most one component dominates the response of each channel at each pixel, demodulation of the filtered component $t_m(\mathbf{x}) = \int_{\mathbb{R}^2} t(\mathbf{x}-\mathbf{p}) g_m(\mathbf{p}) d\mathbf{p}$ can be accomplished with the approximate algorithm

$$\nabla\varphi(\mathbf{x}) \approx \nabla\widehat{\varphi}(\mathbf{x}) = \mathrm{Re}\left[\frac{\nabla t_m(\mathbf{x})}{j t_m(\mathbf{x})}\right], \tag{9}$$

$$a(\mathbf{x}) \approx \widehat{a}(\mathbf{x}) = \left|\frac{t_m(\mathbf{x})}{G_m[\nabla\widehat{\varphi}(\mathbf{x})]}\right|. \tag{10}$$

The approximation errors in the numerator and denominator of (9) are tightly bounded by a *quasi-eigenfunction theorem* [2], [4], [6], [8], [9] involving certain functional norms of $g_m(\mathbf{x})$, $a(\mathbf{x})$, and $\nabla\varphi(\mathbf{x})$. The errors are small, provided that $g_m(\mathbf{x})$ is spatially localized, and $t_m(\mathbf{x})$ is locally coherent. By applying (9) and (10) to the responses of *all* filterbank channels, estimates of the modulating functions of *all* components are produced at *all* pixels.

II. Computation of the Representation

Given an initial phase sample, the component model (1) is determined by the amplitude $a(\mathbf{x})$ and the horizontal and vertical frequencies $\varphi^x(\mathbf{x}) = \frac{\partial}{\partial x}\varphi(\mathbf{x})$ and $\varphi^y(\mathbf{x}) = \frac{\partial}{\partial y}\varphi(\mathbf{x})$. The problem of computing the representation is that of determining, at each pixel, the number of components present and which channel to use in estimating the modulating functions of each component. We introduce an artificial temporal causality relationship between points in the sampled image domain by mapping them to a 1-D lattice according to a path function $\mathcal{O}$: $\mathbf{x} \longmapsto k$, $k \in \mathbb{N}$. This reparamaterization maps the three modulating functions of a component according to $a(\mathbf{x}) \overset{\mathcal{O}}{\longmapsto} a(k)$, $\varphi^x(\mathbf{x}) \overset{\mathcal{O}}{\longmapsto} \varphi^x(k)$, $\varphi^y(\mathbf{x}) \overset{\mathcal{O}}{\longmapsto} \varphi^y(k)$. Suppose two components exist in a region comprising $N+1$ pixels. As the pixels are traversed in order, k varies from 0 to N. Concomitantly, the frequency $\nabla\varphi(k)$ of each component maps out a path in the frequency domain, as depicted in Fig. 3. Let ρ denote continuous arc length along $\mathcal{O}$ and use the notation

$$a'(k) = \frac{\partial}{\partial k}a(k) = \frac{\partial}{\partial\rho}a(\rho)\bigg|_{\rho=k} \tag{11}$$

to indicate the restriction to the discrete 1-D lattice of the derivatives of the modulating functions taken with respect to ρ. Then, $a(k)$ and $a'(k)$ may be expanded in first- and zeroth-order Taylor series, respectively, as follows:

$$a(k+1) = a(k) + a'(k) + \int_k^{k+1} (k+1-\rho)\frac{\partial^2}{\partial\rho^2}a(\rho)d\rho, \tag{12}$$

$$a'(k+1) = a'(k) + \int_k^{k+1} \frac{\partial^2}{\partial\rho^2}a(\rho)d\rho. \tag{13}$$

The components of $\nabla\varphi(k)$ may each expanded in analogous series.

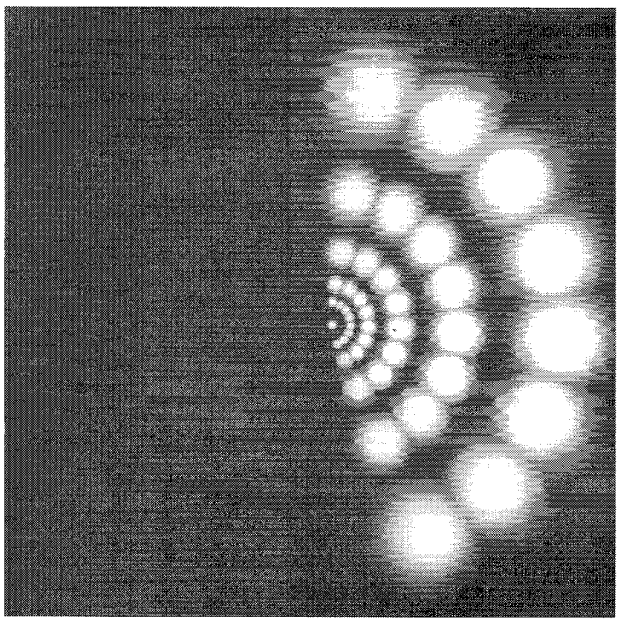


Fig. 2. Frequency domain representation of the filterbank for choices of the design parameters $r_0 = 9.6$ cycles per image, $R = 1.8$, $B = 1$ octave, and $\eta = \frac{1}{2}$. There are 40 filters arranged in a polar wavelet-like tesselation on eight rays with five filters per ray, plus one filter centered at DC. Each of the filters in the figure has been independently scaled for maximum dynamic range in the available grey levels.

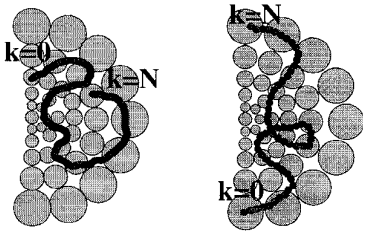


Fig. 3. Frequency tracks of two image components in a region. The index k advances as image pixels are traversed in the spatial domain.

A. Statistical State-Space Component Model

We model $a(\mathbf{x})$ and $\varphi(\mathbf{x})$ as independent, homogeneous, m.s. differentiable random fields and require that $\varphi(\mathbf{x})$ has quadrant symmetry [10]. We write $u_a(k)$, $u_{\varphi_x}(k)$, $u_{\varphi_y}(k)$ for the integral in (12), $v_a(k)$, $v_{\varphi_x}(k)$, $v_{\varphi_y}(k)$ for that in (13), and refer to these six processes collectively as the *modulation accelerations* (MA's). Restating (12) and (13) for all three modulating functions, we obtain the statistical state-space model

$$\begin{bmatrix} a(k+1) \\ a'(k+1) \\ \varphi^x(k+1) \\ \varphi^{x\prime}(k+1) \\ \varphi^y(k+1) \\ \varphi^{y\prime}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a(k) \\ a'(k) \\ \varphi^x(k) \\ \varphi^{x\prime}(k) \\ \varphi^y(k) \\ \varphi^{y\prime}(k) \end{bmatrix} + \begin{bmatrix} u_a(k) \\ v_a(k) \\ u_{\varphi_x}(k) \\ v_{\varphi_x}(k) \\ u_{\varphi_y}(k) \\ v_{\varphi_y}(k) \end{bmatrix} \quad (14)$$

with output vector $[a(k) \quad \varphi^x(k) \quad \varphi^y(k)]^T$. We relate the state-space model output vector to the observations $\widehat{a}(k)$, $\nabla\widehat{\varphi}(k)$ obtained from (9) and (10) by modeling the estimation errors with noise processes $n_a(k)$, $n_{\varphi_x}(k)$, and $n_{\varphi_y}(k)$ called the *measurement noises*

$$\begin{bmatrix} \widehat{a}(k) \\ \widehat{\varphi}^x(k) \\ \widehat{\varphi}^y(k) \end{bmatrix} = \begin{bmatrix} a(k) \\ \varphi^x(k) \\ \varphi^y(k) \end{bmatrix} + \begin{bmatrix} n_a(k) \\ n_{\varphi_x}(k) \\ n_{\varphi_y}(k) \end{bmatrix}. \quad (15)$$

With appropriately specified noise processes, the component model (14) and (15) is statistically valid for *any* image with homogeneous, differentiable amplitude and quadrant symmetric, differentiable phase under the assumption that filterbank isolates components on a localized basis. It follows from homogeneity that the MA's are zero mean. Once specific values are assigned to the noise powers and $\pi(0)$, which is the initial state vector covariance matrix of (14), the model is statistically valid only for specific classes of image components. However, even under a single choice of the parameters, wide ranges of spatial frequencies, and therefore broad classes of textured regions, can be characterized. Furthermore, for locally coherent components, variations in $a(k)$ and $\nabla\varphi(k)$ are necessarily smooth and not too erratic. This implies that the MA noise powers are relatively small, which is consistent with the usual assumptions that the modulating functions have small derivatives or are bandlimited [1], [5]. Therefore, we expect the model (14) to be effective for describing broad classes of locally coherent components.

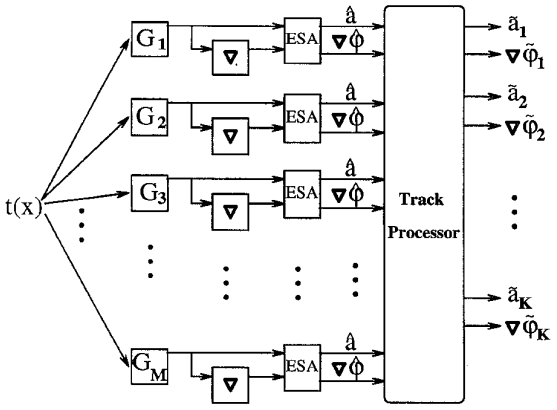


Fig. 4. Block diagram of the approach for computing the multicomponent AM–FM representation of the image $t(\mathbf{x})$.

If $\pi(0)$ and the noise powers are known, determination of the filterbank channel from which to take observations $\widehat{a}(k)$ and $\nabla\widehat{\varphi}(k)$ for any given component is made by following the component frequency domain track, as depicted in Fig. 3, with a track processor. At each pixel along $\mathcal{O}$, this processor predicts $\nabla\widehat{\varphi}(k)$ and $\widehat{a}(k)$, and observations of these quantities are taken from the channel with center frequency closest to the predicted frequencies. Optimal estimates of the modulating functions are then computed from the observations and track history. A block diagram of this approach is shown in Fig. 4, where the modulating functions are estimated for *all* components of a K-component image $t(\mathbf{x})$. The discrete-domain image is processed with an M-channel filterbank, and estimates $\widehat{a}(k)$, $\nabla\widehat{\varphi}(k)$ are produced by each of the M channels at every pixel. From these observations, the track processor formulates optimal estimates $\tilde{a}(k)$, $\nabla\tilde{\varphi}(k)$ for each of the K image components.

Finally, we assume that the correlation structures of $a''(\mathbf{x})$, $\varphi^{x\prime\prime}(\mathbf{x})$, and $\varphi^{y\prime\prime}(\mathbf{x})$ are localized so that they appear impulse-like at the scale of the spatial lattice induced by $\mathcal{O}$. Under this mild assumption, the system modes of (14) corresponding to the amplitude modulation and the horizontal and vertical components of the frequency modulation can be decoupled to yield three separate second-order systems.

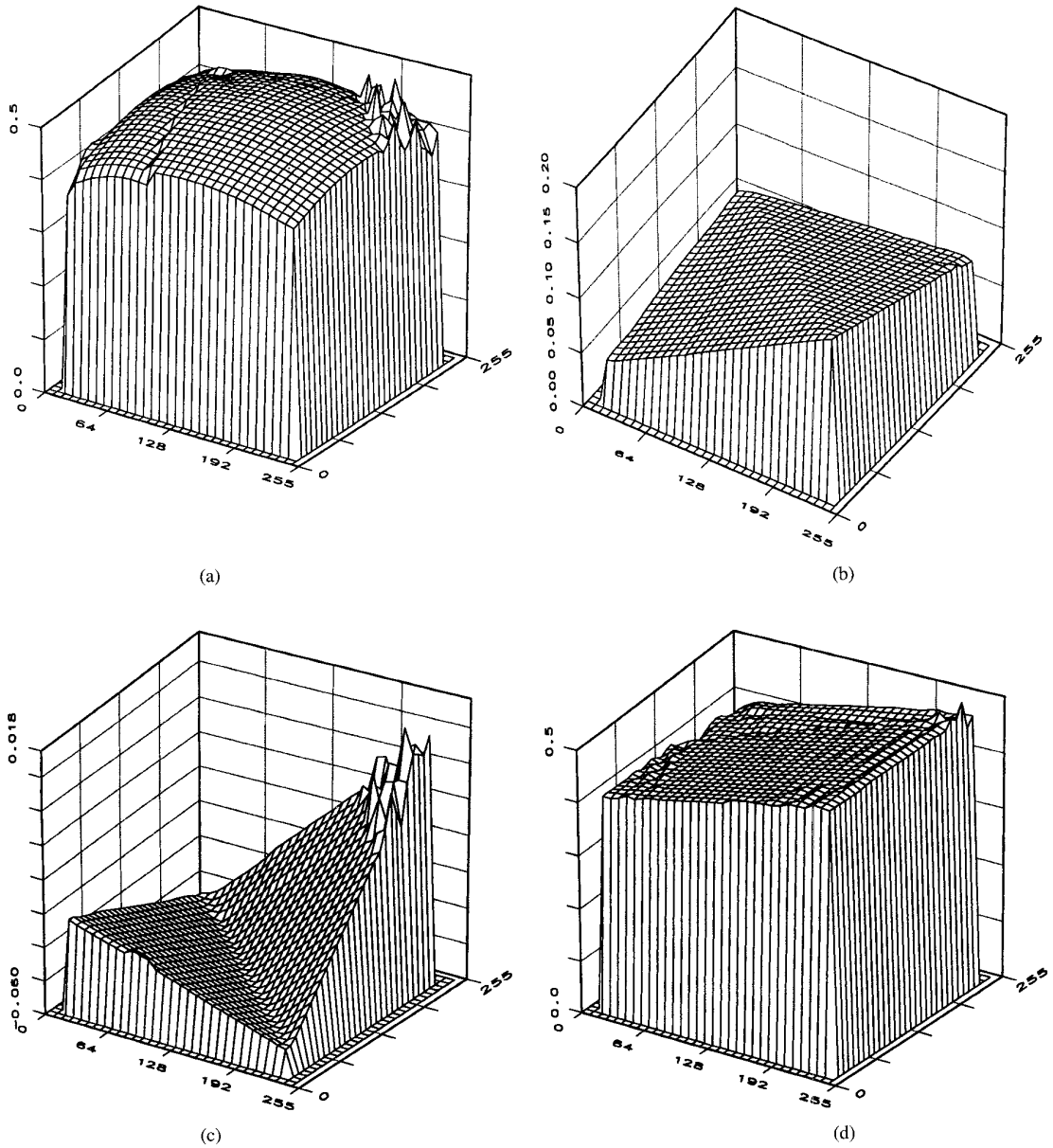


Fig. 5. Multicomponent AM–FM representation and reconstruction of synthetic image. (a) Amplitude estimates for component one. (b) Horizontal frequency estimates for component one. (c) Vertical frequency estimates for component one. (d) Amplitude estimates for component two.

B. Track Processor

We have modeled the estimated modulating functions of a component as noisy observations of an affine function of the state vector of a finite-dimensional linear system driven by uncorrelated noise. Therefore, the MMSE optimal linear track processor involves Kalman filters. The explicit recursive formulation for the optimal amplitude estimates $\tilde{a}(k \mid k)$ is

$$\tilde{a}(k \mid k) = \tilde{a}(k \mid k-1) + \alpha_a(k)(\hat{a}(k) - \tilde{a}(k \mid k-1)), \tag{16}$$

$$\tilde{a}'(k+1 \mid k) = \tilde{a}'(k \mid k-1) + \beta_a(k)(\hat{a}(k) - \tilde{a}(k \mid k-1)), \tag{17}$$

$$\tilde{a}(k+1 \mid k) = \tilde{a}(k \mid k) + \tilde{a}'(k+1 \mid k). \tag{18}$$

Likewise, the optimal frequency estimators are

$$\tilde{\varphi}^x(k \mid k) = \tilde{\varphi}^x(k \mid k-1) + \alpha_{\varphi_x}(k) \times (\hat{\varphi}^x(k) - \tilde{\varphi}^x(k \mid k-1)), \tag{19}$$

$$\tilde{\varphi}^{x\prime}(k+1 \mid k) = \tilde{\varphi}^{x\prime}(k \mid k-1) + \beta_{\varphi_x}(k) \times (\hat{\varphi}^x(k) - \tilde{\varphi}^x(k \mid k-1)), \tag{20}$$

$$\tilde{\varphi}^x(k+1 \mid k) = \tilde{\varphi}^x(k \mid k) + \tilde{\varphi}^{x\prime}(k+1 \mid k) \tag{21}$$

and

$$\tilde{\varphi}^y(k \mid k) = \tilde{\varphi}^y(k \mid k-1) + \alpha_{\varphi_y}(k) \times (\hat{\varphi}^y(k) - \tilde{\varphi}^y(k \mid k-1)), \tag{22}$$

$$\tilde{\varphi}^{y\prime}(k+1 \mid k) = \tilde{\varphi}^{y\prime}(k \mid k-1) + \beta_{\varphi_y}(k) \times (\hat{\varphi}^y(k) - \tilde{\varphi}^y(k \mid k-1)), \tag{23}$$

$$\tilde{\varphi}^y(k+1 \mid k) = \tilde{\varphi}^y(k \mid k) + \tilde{\varphi}^{y\prime}(k+1 \mid k). \tag{24}$$

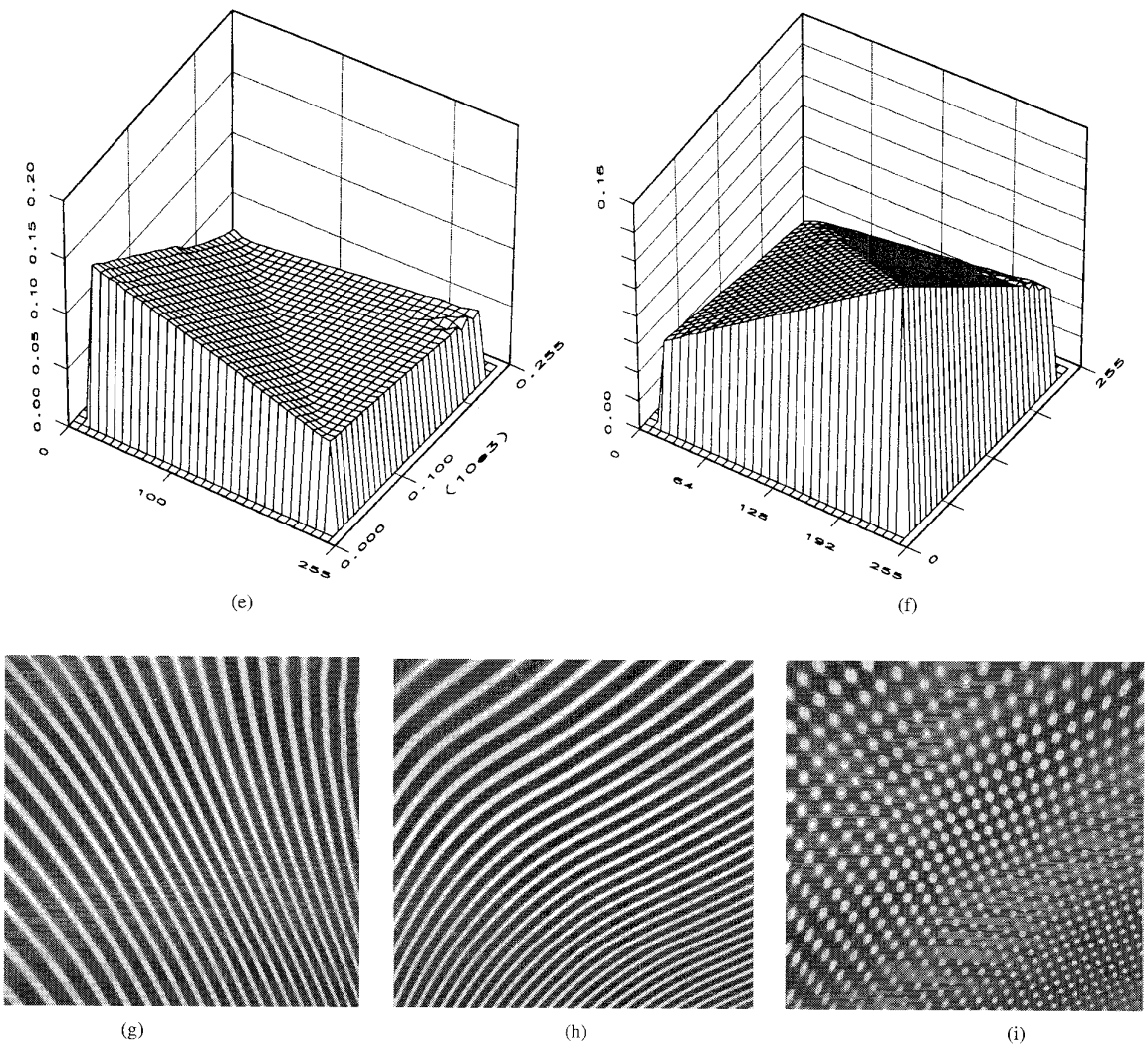


Fig. 5. *(Continued.)* Multicomponent AM–FM representation and reconstruction of the synthetic image: (e) Horizontal frequency estimates for component two; (f) vertical frequency estimates for component two; (g) reconstruction of component one; (h) reconstruction of component two; (i) reconstructed image.

Formulations for the gain sequences $\alpha_{\square}(k)$, $\beta_{\square}(k)$ follow from recursive expressions for the state vector error covariance matrices associated with the decoupled systems. The track processor (16)–(24) is unbiased in that $E[\tilde{a}(k) \quad \tilde{\varphi}^x(k) \quad \tilde{\varphi}^y(k)]^T = [a(k) \quad \varphi^x(k) \quad \varphi^y(k)]^T$ and optimal in the MMSE sense.

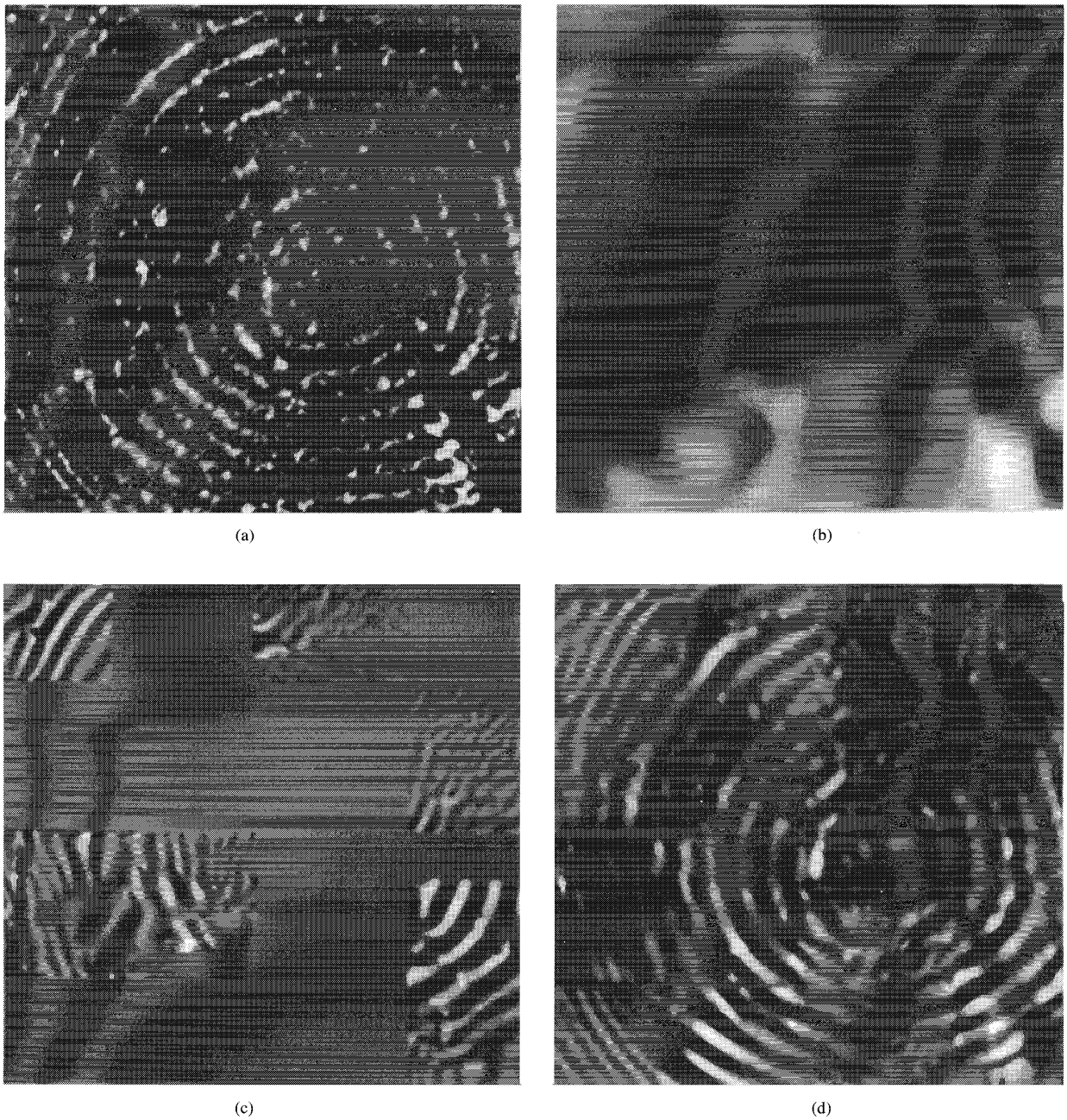
C. New Track Starts

At each point i in $\mathcal{O}$, we start tracks using observations from channels maximizing $Q(n) = \widehat{a}_n(i)|G_n(\nabla\widehat{\varphi}_n(i))|/\max_\Omega |G_n(\Omega)|$, where $\widehat{a}_n(i)$ and $\nabla\widehat{\varphi}_n(i)$ are the estimates obtained from filterbank channel n, provided that these do not associate with an already existing track. Note that the quantity $|G_n(\Omega)|/\max_\Omega|G_n(\Omega)|$ lies between zero and one and increases as Ω moves closer to the center frequency of the channel. Hence, for a given component, tracking will be initialized using the filter with center frequency closest to the frequency of the component, affording improved immunity against out-of-band information through an enhanced SNR.

The amplitude component of the track is initialized by taking $\tilde{a}(0 \mid 0) = \widehat{a}_n(i)$ and $\tilde{a}'(0 \mid 0) = \widehat{a}_n(i) - \widehat{a}_n(i-1)$. The frequency components are initialized similarly, and the initial predictions are computed from the transformation equation $\tilde{X}(k+1 \mid k) = A(k)\tilde{X}(k|k)$. Typically, for some small number L, new tracks are started at each pixel using observations from the L channels maximizing $Q(n)$ whose observations do not associate with an existing track. Under the assumption that at most one component dominates the local frequency spectrum of each channel at each pixel, in an image region where K components are present this scheme guarantees that *all* components will be tracked by the $\lceil K/L \rceil$th pixel of the region.

III. Reconstruction

Reconstruction of the image from the representation (3) is a difficult, ill-posed inverse problem. The phase $\varphi_k(\mathbf{x})$ of the kth component could be recovered from perfect continuous-domain frequency estimates $\nabla\tilde{\varphi}_k(\mathbf{x})$ by integrating along arbitrary paths. However, due to estimation errors arising from cross-component interference and the approximation error inherent in (9), the estimated phase gradient field is not, in practice, conservative. Therefore, summing frequency estimates around arbitrary closed paths will generally not yield zero. Since estimates of both components of the phase gradient are available at every pixel, the phase reconstruction problem is, in fact, overdetermined. While various interpolation schemes



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

Fig. 6. Multicomponent AM–FM representation and reconstruction of the complicated, natural multipartite Brodatz texture image *tree*. (a) Tree image. (b) Lowpass component extracted by linear filtering. (c) reconstructions of five tracked AM–FM components. (d) Reconstruction of the image from 41 AM–FM components and the lowpass component. Most pixels in the image are covered by *three or fewer* reconstructed AM–FM components.

could be employed, we have found that for components supported in a rectangular region, it is often satisfactory to reconstruct the phase using an asymmetric difference interpretation of the derivative according to

$$\tilde{\varphi}_k(m,n) = \frac{1}{2}[\tilde{\varphi}_k(m-1,n) + \tilde{\varphi}_k^x(m-1,n) + \tilde{\varphi}_k(m,n-1) + \tilde{\varphi}_k^y(m,n-1)]. \quad (25)$$

The phase is reconstructed along the top row of the component by iterating

$$\tilde{\varphi}_k(m,n_0) = \tilde{\varphi}_k(m-1,n_0) + \tilde{\varphi}_k^x(m-1,n_0) \quad (26)$$

and an analogous scheme is used along the first column. Phase reconstruction is initialized by saving an estimated initial phase sample

$$\tilde{\varphi}_k(m_0,n_0) = \arctan\left\{\frac{\mathrm{Im}[t_i(m_0,n_0)]}{\mathrm{Re}[t_i(m_0,n_0)]}\right\} \quad (27)$$

during computation of the representation (3), where tracking of the kth component was begun at pixel (m_0,n_0) using the amplitude and frequency observations obtained from filterbank channel i, and $t_i(m,n)$ is the response of the ith channel filter.

With this approach, the deleterious effects of estimation errors are cumulative. We improve the robustness of the technique by saving phase estimates $\tilde{\varphi}_k(m,n)$ on a rectangular grid and reconstructing the phase on each rectangle independently. Each component is then reconstructed using (1), and the image is subsequently recovered from (7).

IV. Examples

The multicomponent paradigm depicted in Fig. 4 was applied to compute the representation (3) of the 256×256 two-component image of Fig. 1. The track processor correctly identified the presence of two components. Amplitude estimates $\tilde{a}_1(m,n)$ for component one are shown in Fig. 5(a), whereas frequency estimates $\tilde{\varphi}_1^x(m,n)$ and $\tilde{\varphi}_1^y(m,n)$ are given in Fig. 5(b) and (c), respectively. These quantities are in near-perfect agreement with the true values. The small region of oscillatory behavior visible in the vertical frequency estimates corresponds to the top right-most portion of the component as depicted in Fig. 1(b), where the vertical frequency is nearly zero, and occurs because, at very low frequencies, the distinction between which image features should be interpreted as amplitude modulation as opposed to frequency modulation becomes unclear. Note that this oscillatory behavior also propagates into the amplitude estimates as a consequence of (10). Amplitude and frequency estimates $\tilde{a}_2(m,n)$, $\tilde{\varphi}_2^x(m,n)$, and $\tilde{\varphi}_2^y(m,n)$ for component two are shown in Fig. 5(d)–(f). Once again, these estimates are in near-perfect agreement with the true values. Collectively, Fig. 5(a)–(f) constitute the multicomponent AM–FM representation (3) of the image. Note how smooth the AM–FM representation is, despite the fact that there are rapid variations in the image. Significant compression of the estimated quantities could be achieved, e.g., through linear predictive coding.

The individual components were reconstructed on a 32×32 pixel grid of estimated phase samples (27) using the algorithm (25), (26). The reconstructions of components one and two are shown in Fig. 5(g) and (h), respectively. Visually, the reconstructed components are virtually indistinguishable from the true values. Reconstruction of the image using (7) is shown in Fig. 5(i) and is virtually indistinguishable from the true values in Fig. 1(a).

As a final preliminary validation of the power of this new representation for analyzing and representing natural images, we computed a multicomponent AM–FM representation for the tree image of Fig. 6(a). This complicated, nonstationary multipartite image presents significant challenges since the number of components, the sizes and shapes of their regions of support, and their statistical characteristics are unknown. The filterbank of Fig. 2 was again used, but in this case, the assumption that each channel is dominated by at most one component at each pixel was almost certainly violated. In many parts of the image, there appear to be many components and harmonics closely spaced in frequency, giving rise to errors in algorithm (9), (10). Furthermore, there is no way to determine the optimal Kalman gain sequences, and hence, the computed representation is suboptimal.

As we have noted, very low-frequency information is not particularly well suited to computed AM–FM modeling due to the ambiguity inherent in determining which features of the low-frequency structure should be interpreted as AM as opposed to FM. Hence, we began by extracting the low-frequency component shown in Fig. 6(b) by linear filtering prior to computation of the representation. Next, we heuristically divided the image into rectangular regions and applied the approach of Fig. 4 to each one. Many tracks were generated in each region, and we hand selected 41 of them. Reconstruction was performed on 4×4 pixel grids, and five of the reconstructed components are shown in Fig. 6(c). Summing all of the reconstructed components with the low-frequency component of Fig. 6(b), we obtained the excellent reconstruction shown in Fig. 6(d). Most pixels in the image are covered by *three or fewer* reconstructed AM–FM components. It is indeed compelling that, despite being suboptimal, the computed AM–FM representation has clearly succeeded in capturing the essential structure, features, and information content of the image.

V. Conclusion

The multicomponent AM–FM representation is a powerful, important new emerging technique for modeling, analysis, and representation in a general image processing framework. Multipartite images are characterized as sums of locally coherent, nonlinear AM–FM functions capable of effectively capturing nonstationarities, where important visual and perceptual cues are often manifest. Using a statistical state-space component model, we developed practical algorithms for computing the representation, as well as for recovering an image from its computed representation. The approach bears similarities to certain processing known to occur in biological vision systems in that the information content of an image is represented by smoothly varying modulations occurring in a few frequency and orientation selective channels. Under the assumptions that the filterbank and track processor gain sequences are correctly designed, the computed representation is both unbiased and optimal in the mean squared error sense. Tracking of all components is also guaranteed. We computed multicomponent AM–FM representations for two multipartite images and obtained reconstructions in remarkable agreement with the original images. Important future work remaining in this area will address the extremely difficult problems in treating complicated natural images, which may contain many components closely spaced in frequency and supported on irregularly shaped regions.

References

[1] P. Maragos and A. C. Bovik, "Image demodulation using multidimensional energy separation," *J. Opt. Soc. Amer. A*, vol. 12, no. 9, pp. 1867–1876, Sept. 1995.

[2] J. P. Havlicek and A. C. Bovik, "Multi-component AM-FM image models and wavelet-based demodulation with component tracking," in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, Nov. 13–16, 1994, pp. I41–I45.

[3] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with applications to speech analysis," *IEEE. Trans. Signal Processing*, vol. 41, no. 10, pp. 3024–3051, Oct. 1993.

[4] J. P. Havlicek, A. C. Bovik, and P. Maragos, "Modulation models for image processing and wavelet-based image demodulation," in *Proc. 26th IEEE Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 26–28, 1992, pp. 805–810.

[5] P. Maragos, A. C. Bovik, and T. F. Quatieri, "A multidimensional energy operator for image processing," in *Proc. SPIE Symp. Visual Commun. Image Processing*, Boston, MA, Nov. 16–18, 1992, pp. 177–186.

[6] J. P. Havlicek and A. C. Bovik, "AM-FM models, the analytic image, and nonlinear demodulation techniques," Tech. Rept., Cent. Vision and Image Sciences, Univ. of Texas at Austin, no. TR-95-001, Mar. 1995.

[7] L. Cohen and C. Lee, "Instantaneous bandwidth for signals and spectrogram," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 2451–2454.

[8] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo, "Localized measurement of emergent image frequencies by Gabor wavelets," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 691–712, Mar. 1992.

[9] J. P. Havlicek, A. C. Bovik, M. D. Desai, and D. S. Harding, "The discrete quasi-eigenfunction approximation," in *Proc. Int. Conf. Digital Signal Processing*, Limassol, Cyprus, June 26–28, 1995.

[10] E. Vanmarcke, *Random Fields, Analysis and Synthesis*. Cambridge, MA: MIT Press, 1983.