

"Computation of the Multi-Component AM-FM
Image Representation"

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

UT-CVIS-TR-96-001

May, 1996



CENTER FOR VISION AND IMAGE SCIENCES

The University of Texas at Austin, Mezes Hall 330, Austin, TX 78712

Computation of the Multi-Component AM-FM Image Representation

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

Laboratory for Vision Systems

University of Texas, Austin, TX 78712-1084

Tech. Rept. TR-96-001

Center for Vision and Image Sciences, The University of Texas at Austin

May 7, 1996

Abstract – In this report we present very general, powerful multi-component AM-FM image models capable of efficiently representing complicated nonstationary multi-partite images. The nonstationarities in such images, which often contribute significantly to visual perception and interpretation, are difficult to capture using stationary frequency techniques.

Highly localized nonlinear operators are developed for simultaneously estimating the amplitude and frequency modulating functions associated with each of the multiple components on a pointwise basis. Since the demodulation algorithms are nonlinear, the components must be isolated from one another prior to demodulation. This is accomplished with a multi-band bank of optimally spatio-spectrally localized Gabor filters in a wavelet-like tessellation. We introduce a statistical state-space image component model, and use it to develop optimal filters for tracking the estimated modulating functions of each image component across the filterbank channel responses. We develop practical techniques both for computing the multi-component AM-FM representation of an image and for recovering the image from the representation.

This research 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 number F49620-93-1-0307.

Contents

1	Introduction	4
1.1	AM-FM Modeling	7
2	Demodulation of Multiple Components	12
3	Computation of the Representation	14
3.1	Statistical State-Space Component Model	15
3.2	Specializing the Model	19
3.2.1	First-order Moments of Modulation Accelerations	19
3.2.2	Second-order Moments of Modulation Accelerations	19
3.2.3	Measurement Noise	22
3.2.4	State Vector Covariance Matrix	24
3.2.5	Summary of Assumptions	24
3.3	Track Processor	30
3.4	New Track Starts	33
4	Reconstruction	34
5	Examples	36
6	Conclusion & Future Work	40

List of Figures

1	Quantities used in defining the bandwidth B of a Gabor filter. (a) Radial octave bandwidth. The curve shows the filter evaluated on a line from the frequency origin through it's center frequency r_m . The filter magnitude response is at a fraction η of peak response at the radial frequencies r_1 and r_2 . (b) Orientation bandwidth. The circle with center r_m is the η -peak contour of the filter in the $\mathbf{\Omega} = [u, v]^T$ plane.	10
2	Frequency domain representation of the filter bank 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 tessellation on eight rays with five filters per ray, plus one filter centered at DC. Each of the 41 filters in the figure has been independently scaled for maximum dynamic range in the available grey levels.	11
3	Nonstationary multi-component synthetic image showing one possible decomposition into highly locally coherent components. (a) Composite image. (b) Component zero. (c) Component one. (d) Histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for the composite image. (e) Histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for component zero. (f) Histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for component one. Note the difference in the abscissa scale between the (d) part of the figure and the (e) and (f) parts, which clearly indicates the multi-partite nature of this image.	13
4	Frequency tracks of two image components in a region. The index k advances as we traverse the image pixels in the spatial domain.	15
5	Real parts of 16 image components generated from the statistical state-space model using values $\sigma_{u_a} = 5 \times 10^{-8}$ and $\sigma_{u_{\varphi_x}} = \sigma_{u_{\varphi_y}} = 10^{-7}$ (see text for values of $\boldsymbol{\pi}(0)$).	27
6	Path function $\mathcal{O} : \mathbf{x} \longrightarrow k$ used to generate the sixteen image components shown in Figure 5, as given by Equation (61).	28
7	Block diagram of the tracked multi-component paradigm.	29
8	Multi-component AM-FM representation and reconstruction of a synthetic image. (a) Nonstationary two-component image. (b) True values for component one. (c) True values for component two. (d) Amplitude estimates for component one. (e) Horizontal frequency estimates for component one. (f) Vertical frequency estimates for component one. (g) Amplitude estimates for component two. (h) Horizontal frequency estimates for component two. (i) Vertical frequency estimates for component two. (j) Reconstruction of component one. (k) Reconstruction of component two. (l) Reconstructed image.	38
9	Reconstructions of 20 AM-FM components of the tree image from their amplitude and frequency estimates.	41
10	Reconstructions of 16 AM-FM components of the tree image from their amplitude and frequency estimates.	42
11	Reconstructions of 5 AM-FM components of the tree image from their amplitude and frequency estimates.	43

- 12 Reconstruction of the complicated natural multi-partite Brodatz texture image *tree* from 41 computed AM-FM components and a low-frequency component. (a) Tree image. (b) Low-pass component extracted by linear filtering. (c) Reconstruction from 41 AM-FM components and the low-pass component. 43

1 Introduction

In this report, we develop a new multi-component AM-FM representation for images, present an algorithm for computing the representation, and demonstrate reconstruction of images from their computed multi-component AM-FM representations. To compute the representation, an image is considered as a sum of components each modeled by locally coherent, complex-valued AM-FM functions of the form

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

where in the general n -dimensional case $\mathbf{x} = (x_1, x_2, \dots, x_n)$, $t : \mathbb{R}^n \rightarrow \mathbb{C}$, $a : \mathbb{R}^n \rightarrow [0, \infty)$, and $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$, or as a sum of locally coherent real-valued components

$$s(\mathbf{x}) = a(\mathbf{x}) \cos [\varphi(\mathbf{x})]. \quad (2)$$

By *locally coherent*, we mean that the amplitude and frequency modulating signals $a(\mathbf{x})$ and $\nabla\varphi(\mathbf{x})$ vary smoothly in the sense of having tightly bounded first-order Sobolev norms. Although the Sobolev norm is a *global* measure of smoothness, it is true in general that the smaller the Sobolev norms, the more locally coherent the component (local coherency will be quantified on a pointwise basis in Section 2). Some advantages of the multi-component AM-FM representation may not be realized with fractal, self-similar, or extraordinarily discontinuous images, which are termed *incoherent* in this sense.

In terms of *image representation*, an important question that arises naturally is *for what class of images is the multi-component AM-FM representation more efficient, more natural, or more convenient than the standard Fourier frequency representation?* A related question that arises under the auspices of *image analysis* is *for what class of images is the multi-component AM-FM representation more intuitive, more illuminating, or better suited for revealing the structure of an image than the standard Fourier frequency representation?*

For an image comprising N^2 pixels, the 2D DFT represents the image in terms of its projection onto N^2 complex sinusoidal gratings, each supported over the entire image. If the image is, in fact, composed of a small number of such gratings, then *most* of the energy

will be concentrated in a small number of the DFT frequency coefficients. In this case, the DFT representation will be *efficient* in the sense that only a small number of sinusoidal components are necessary to represent the image, and will also reveal the inherent structure of the image.

However, the complex sinusoidal grating, which is the kernel of the DFT, is also the simplest AM-FM function: both amplitude and frequency modulations are constant over the entire domain. Therefore, multi-component AM-FM representations *always* exist that are *at least* as efficient as the DFT representation, and that are *at least* as suitable for revealing structure. Indeed, we expect that there may be nonstationary features in an image — features that are better represented in terms of quasi-sinusoidal gratings whose amplitudes and frequencies are allowed to vary across the domain. This idea forms the core impetus motivating multi-component AM-FM modeling. If such components are present in an image, then the DFT will, in general, admit many nonzero frequency coefficients, giving a representation in terms of a linear combination of a large number of kernel functions. If a component is present in one region of the image with a spatial frequency of Ω , and this frequency changes slightly to $\Omega + \epsilon$ in an adjacent region, then DFT kernel functions (*i.e.* complex sinusoidal gratings) of *both* frequencies must be present over the *entire domain* in the DFT representation. In regions where either frequency is not directly present in the image, additional gratings of related frequencies must appear in the DFT representation to *cancel* the unwanted frequencies (destructive interference). This feature of the 2D DFT decreases the *efficiency* of the representation, and also tends to obscure the inherent structure of the image from an analysis standpoint. Clearly, relaxing the kernel functions to allow nonstationary, spatially varying amplitude and frequency modulations ameliorates these shortcomings in the context of analyzing, interpreting, and representing nonstationary images. For an image with nonstationary features and structure, multi-component AM-FM representations will *always* exist that are *more efficient* than the DFT representation and that *better* reveal the image structure. In the event that nonstationarities are *not* present, multi-component AM-FM representations will *always* exist that are *at least* as efficient as the DFT representation, and *at least* as well suited for revealing the structure (we reiterate that the DFT representation *is*, in fact, a multi-component AM-FM representation where

each component is restricted to be monochromatic; *e.g.* to have *constant* modulating functions). Indeed, the multi-component AM-FM representation is an exciting, *new* framework in which to treat images of *all classes*, although its maximum utility is realized when applied to images which exhibit the mildly restraining characteristic of *local coherency*.

Many images may be characterized in terms of multiple nonstationary components, some of which are supported only on certain (potentially irregularly shaped) regions of the image. Within the theoretical framework of the multi-component AM-FM representation, there is no inherent difficulty in accommodating such regionally supported components. They may be modeled with AM-FM functions with amplitude modulations which become zero in regions where the component is not supported. However, the development of practical techniques for computing the multi-component AM-FM representations of such images is a difficult problem. The development of a fully viable and satisfactory practical technique for assimilating the estimated amplitude and frequency modulating functions arising from large numbers of components closely spaced in frequency and supported on irregularly shaped subregions of an image is ongoing. Nevertheless, in Section 5 we compute the multi-component AM-FM representation of a complicated natural Brodatz texture image. It is both exciting and compelling that, when we reconstruct the image from the computed representation, it immediately becomes apparent that the essential structure, features, and information content of the 256×256 image have been successfully captured using only 41 regionally-supported AM-FM components. In fact, most pixels of the image are covered by *three or fewer* computed AM-FM components. This stunning preliminary result validates the power and importance of the multi-component AM-FM model as a new, emerging technique for modeling, analysis, representation, and interpretation in a completely general image processing framework.

In an earlier related report, we introduced the concept of the *analytic image*, developed nonlinear techniques for estimating the AM and FM modulating functions of a single image component, and presented an analysis paradigm called *dominant component analysis* [1]. While the emphasis of the present report is the development of the more powerful multi-component AM-FM representation, there will necessarily be a degree of overlap with the previous report, which is required in order for this report to stand alone.

We are interested in analyzing and representing images containing regions of surface

texture or patterned markings that, while nonstationary and globally wideband, can be effectively modeled as being locally narrowband, in the sense that on a spatially local basis the image spectrum is dominated by narrow distributions of frequencies centered about a single, or small number of *emergent* frequencies. If the locally emergent frequencies do not vary too wildly or discontinuously on a spatially local scale, then such regions are described as being *locally coherent*. The development of techniques for characterizing and representing such images in a way that facilitates and simplifies analysis in terms of the locally occurring nonstationarities, which manifest important structural and perception cues, remains an important open problem. Indeed, the characterization and analysis of images in terms of their instantaneous frequency content on a spatially local basis is fundamental to an increasing variety of processing techniques in machine vision, image processing, and computational vision, including analysis, segmentation and modeling of texture [2–6], shape from texture [7], and stereopsis [8].

Instantaneous frequency estimation based on AM-FM modeling techniques has been the focus of significant recent research, and the efficacy of such techniques for analyzing and characterizing locally coherent nonstationary signal and image components has been well established [9–23]. In 1D, AM-FM models in terms of non-negative instantaneous frequency can be efficiently computed using the Teager-Kaiser Energy Operator [13–15, 17, 19, 22]. In terms of analyzing and interpreting the structure, information content, and origin of certain important signals (*e.g.* speech signals), the inherent ability of AM-FM models to capture local nonstationarities offers significant advantages over traditional time-frequency distributions [24–31], and high-quality reconstruction of a 1D signal from the AM-FM model has been demonstrated recently [9]. A multi-dimensional version of the Teager-Kaiser operator has been used to compute AM-FM models of images whose instantaneous frequencies are constrained to the first quadrant of the spatial frequency plane [10, 20, 21], and related demodulation algorithms have been applied to more general classes of images [1, 11, 12, 18].

1.1 AM-FM Modeling

AM-FM modeling is most useful when the images of interest may be accurately modeled as locally coherent AM-FM functions, or as sums of a few locally coherent func-

tions. In (1),(2), note that $t(\mathbf{x})$ and $s(\mathbf{x})$ are uniquely related through $s(\mathbf{x}) = \text{Re}[t(\mathbf{x})]$ and $t(\mathbf{x}) = s(\mathbf{x}) + j\mathcal{H}[s(\mathbf{x})]$, where $\mathcal{H}[\cdot]$ indicates the multidimensional Hilbert transform acting in the direction of the unit vector \mathbf{e}_i , defined by [1,31–33]

$$\mathcal{H}[s(\mathbf{x})] = \frac{1}{\pi} \int_{\mathbb{R}} s(\mathbf{x} - y\mathbf{e}_i) \frac{dy}{y} \quad (3)$$

$$= s(\mathbf{x}) * \frac{\delta(\mathbf{x}^T \mathbf{e}_j)}{\pi \mathbf{x}^T \mathbf{e}_i}, \quad (4)$$

where the integral is interpreted as a Cauchy principle value, \mathbf{e}_j is a unit vector orthogonal to \mathbf{e}_i , and δ is the Dirac delta. If $w(\mathbf{x}) = \mathcal{H}[s(\mathbf{x})]$, then the Fourier transforms of $w(\mathbf{x})$ and $s(\mathbf{x})$ are related by

$$W(\boldsymbol{\Omega}) = \mathcal{F}[w(\mathbf{x})] = -j \text{sgn}[\boldsymbol{\Omega}^T \mathbf{e}_i] S(\boldsymbol{\Omega}). \quad (5)$$

We always take the transform in the $\mathbf{e}_x = [1, 0]^T$ direction, which means that the spectrum of the complex image

$$t(\mathbf{x}) = s(\mathbf{x}) + j\mathcal{H}[s(\mathbf{x})] \quad (6)$$

is supported only in quadrants I and IV of the $\boldsymbol{\Omega} = [u, v]^T$ frequency plane. We refer to $t(\mathbf{x})$ as the *analytic image* associated with the real-valued image $s(\mathbf{x})$ [1]. Note that in $\nabla\varphi(\mathbf{x})$ we have an unambiguous definition for the *instantaneous frequencies* of $t(\mathbf{x})$. By the instantaneous frequencies of $s(\mathbf{x})$, we shall mean the (unambiguously defined) instantaneous frequencies of its associated analytic image.

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

$$a(\mathbf{x}) = |t(\mathbf{x})|, \quad (7)$$

$$\nabla\varphi(\mathbf{x}) = \text{Re} \left[\frac{\nabla t(\mathbf{x})}{jt(\mathbf{x})} \right], \quad (8)$$

which is *exact* for *any* general n -dimensional complex valued AM-FM component [12, 18]. For a multi-component image modeled as the real part of the sum

$$t(\mathbf{x}) = \sum_{i=1}^K a_i(\mathbf{x}) \exp[j\varphi_i(\mathbf{x})], \quad (9)$$

the nonlinear demodulation algorithm (7),(8) suffers from cross-component interference between the multiple image components. Therefore, prior to demodulation, it is necessary to separate the individual components from one another. We accomplish this using a bank of multiband filters. The filters must be spectrally localized to prevent interference between components, but also spatially localized to capture nonstationarities in the locally narrow-band components, which in general may be globally broadband.

The design of such a filterbank using a wavelet-like tessellation of Gabor functions, which optimally realize the uncertainty principle lower bound on conjoint spatio-spectral localization, can be found elsewhere [1, 6, 12, 34]. The isotropic unity L^2 -norm baseband filter is $h(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{4}\mathbf{x}^T\mathbf{x}\right]$. Upon adding scaling and frequency modulation (translation) while maintaining unity L^2 -norm and aspect ratio, we obtain the scaled, translated channel filter with center frequency $\mathbf{\Omega}_m$:

$$g_m(\mathbf{x}) = \frac{1}{\sigma_m \sqrt{2\pi}} \exp\left[-\frac{1}{4\sigma_m^2}\mathbf{x}^T\mathbf{x}\right] \exp\left[j2\pi\mathbf{\Omega}_m^T\mathbf{x}\right]. \quad (10)$$

In the frequency domain, the filter is the Gaussian

$$G_m(\mathbf{\Omega}) = \mathcal{F}[g_m(\mathbf{x})] = 2\sigma_m \sqrt{2\pi} \exp\left[-4\pi^2\sigma_m^2(\mathbf{\Omega} - \mathbf{\Omega}_m)^T(\mathbf{\Omega} - \mathbf{\Omega}_m)\right], \quad (11)$$

with radial center frequency $r_m = |\mathbf{\Omega}_m|$ and orientation $\theta_m = \arg[\mathbf{\Omega}_m]$. Figure 1 shows the quantities used to define the bandwidth of the filter. The circle in Figure 1(b) is the η -peak contour of the filter. A line running from the frequency origin through the filter center intercepts the η -peak contour at radial frequencies $r_1 < r_2$, given by $r_1 = \sigma_m - \sqrt{-\ln \eta}$ and $r_2 = \sigma_m + \sqrt{-\ln \eta}$. A section of the filter evaluated along this line is shown in Figure 1(a). The η -peak radial octave bandwidth is defined by $B = \log_2[r_2/r_1]$, which is a design parameter of the filter bank. The orientation bandwidth is the angle between two lines tangent to the η -peak contour and passing through the frequency origin. It is given by $\Theta = 2 \arctan \sqrt{\gamma}$, where $\gamma = (2^B - 1)^2 / (2^B + 1)^2$.

Since the spectra of analytic images are supported only in the right frequency half-plane, we arrange the filter center frequencies in the right half-plane along rays such that any group of four adjacent filters intersect precisely where each is at a fraction η of peak response. With

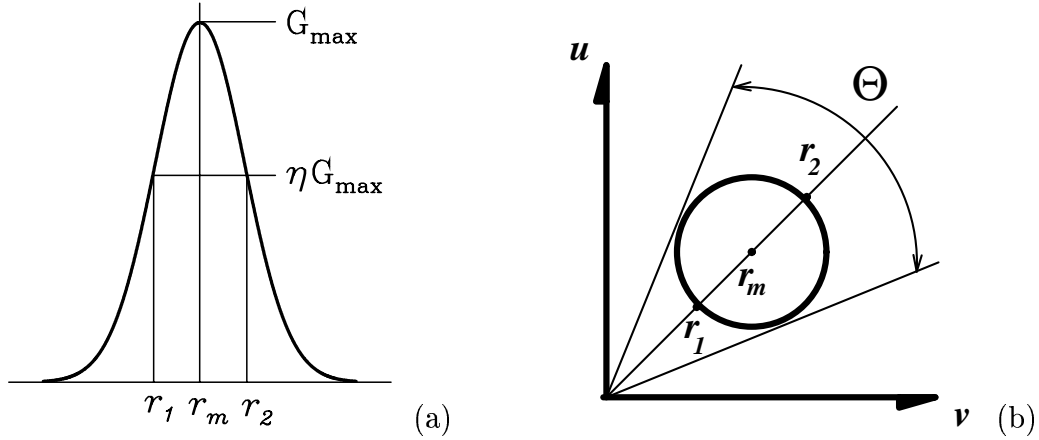


Figure 1: Quantities used in defining the bandwidth B of a Gabor filter. (a) Radial octave bandwidth. The curve shows the filter evaluated on a line from the frequency origin through it's center frequency r_m . The filter magnitude response is at a fraction η of peak response at the radial frequencies r_1 and r_2 . (b) Orientation bandwidth. The circle with center r_m is the η -peak contour of the filter in the $\Omega = [u, v]^T$ plane.

this dense spacing, every point in the right half-plane is covered by a filter responding at η -peak or higher. Along each ray, the filter radial center frequencies progress geometrically with a common ratio R , which is a design parameter of the filterbank. The radial center frequency of the first filter on each ray, r_0 , is also a design parameter. The angular spacing between rays is given by $\Lambda = 2 \arcsin \left[(4R)^{-\frac{1}{2}} \{ (R^2 + 1)(\gamma - 1) + 2R(\gamma + 1) \}^{\frac{1}{2}} \right]$.

The four free design parameters r_0 , R , η , and B completely specify the filterbank. For the examples in this report, we specified the filter bank by $r_0 = 9.6$ cycles per image, $R = 1.8$, $B = 1$ octave, and $\eta = \frac{1}{2}$. In this case, $\gamma = \frac{1}{9}$ and $\Theta \approx 38.9424^\circ$. Figure 2 shows the entire filter bank in the frequency domain for these parameter choices. The frequency domain coordinates of the figure are such that the first quadrant is located in the lower right portion of the figure. Under this convention, a vector from the frequency origin to the spectral support of a sinusoidal grating points *in the direction of propagation* of the grating. There are 40 filters arranged on eight rays spaced equally at angles of $\Lambda = 20.6418^\circ$, with five filters per ray, plus one filter at DC. Each filter in the figure has been *independently* scaled for maximum dynamic range of grey scales.

Assuming that the filterbank has been properly designed, so that at most one component

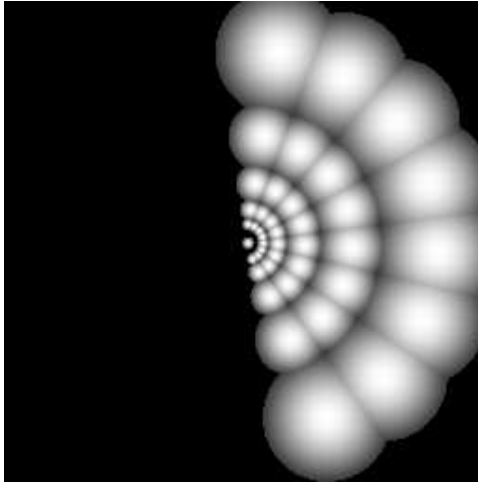


Figure 2: Frequency domain representation of the filter bank for choices of the design parameters $r_0 = 9.6$ cycles per image, $R = 1.8$, $B = 1$ octave, and $\eta = \frac{1}{2}$. There are 40 filters arranged in a polar wavelet-like tessellation on eight rays with five filters per ray, plus one filter centered at DC. Each of the 41 filters in the figure has been independently scaled for maximum dynamic range in the available grey levels.

dominates the response of each channel filter at each pixel, demodulation of the filtered component

$$t_m(\mathbf{x}) = \int_{\mathbb{R}^n} t(\mathbf{x} - \mathbf{p}) g_m(\mathbf{p}) d\mathbf{p} \quad (12)$$

can be accomplished using the approximate algorithm

$$\nabla \varphi(\mathbf{x}) \approx \nabla \hat{\varphi}(\mathbf{x}) = \text{Re} \left[\frac{\nabla t_m(\mathbf{x})}{j t_m(\mathbf{x})} \right], \quad (13)$$

$$a(\mathbf{x}) \approx \hat{a}(\mathbf{x}) = \left| \frac{t_m(\mathbf{x})}{G_m[\nabla \hat{\varphi}(\mathbf{x})]} \right|. \quad (14)$$

In deriving (13),(14), we make use of a *quasi-eigenfunction approximation* [1, 12, 18, 34–37] which tightly bounds the errors in the numerator and denominator of (13) by certain functional norms of $g_m(\mathbf{x})$, $a(\mathbf{x})$, and $\nabla \varphi(\mathbf{x})$. The approximation errors are generally negligible provided that $g_m(\mathbf{x})$ is spatially localized and that the components of $t(\mathbf{x})$ are locally coherent. The algorithm may not work well for components which are everywhere highly incoherent, however.

2 Demodulation of Multiple Components

The problem of analyzing a complicated, nonstationary real-valued image in terms of the model (9) is ill-posed, since there are infinitely many unique sets of modulating functions for which the real part of the model is exactly equal to the image. Given a decomposition of the image into components, the Hilbert transform complex extension technique uniquely determines the imaginary part and instantaneous frequency of each component. However, the question of how the components themselves should be chosen still remains ambiguous. Indeed, we must reiterate that *any* image may be interpreted as comprising a single-component, although such interpretation may not in general admit a smooth AM-FM representation. Hence we begin this section by clarifying, as far as possible, what is meant by the term *multi-partite image*.

For applications in image processing and computational vision, the most useful interpretations are generally those involving modulating functions that are as smooth as possible. This corresponds to a decomposition of the image into components exhibiting as great a degree of local coherency as possible, and which are hence amenable to spatially efficient representation. Therefore, we favor decompositions with a small number of nonstationary components that are as locally quasi-monochromatic as possible. Local coherency may be quantified on a pointwise basis by the instantaneous bandwidth,

$$B(\mathbf{x}) = \left\{ \text{Im} \left[\frac{\nabla t(\mathbf{x})}{jt(\mathbf{x})} \right] \right\}^2 = \frac{|\nabla a(\mathbf{x})|^2}{|a(\mathbf{x})|^2}, \quad (15)$$

the 1D analog of which has been treated by Cohen [24, 26, 27]. The form of (15) should be compared to (8) and (13). For non-monochromatic images, a distribution of instantaneous frequencies is present about the dominant (or emergent) frequencies at every point in the domain of the image. $B(\mathbf{x})$ characterizes the tightness of the spread of this distribution [1, 12, 18, 24, 26, 27]. If, in a certain image region, we can find a multi-component decomposition for which the instantaneous bandwidths of the individual components are appreciably smaller than the instantaneous bandwidth of the composite image in the region, then we say that the image is multi-component in that region. As an example, consider the nonstationary image shown in Figure 3 (a). One possible decomposition of the image into locally coherent

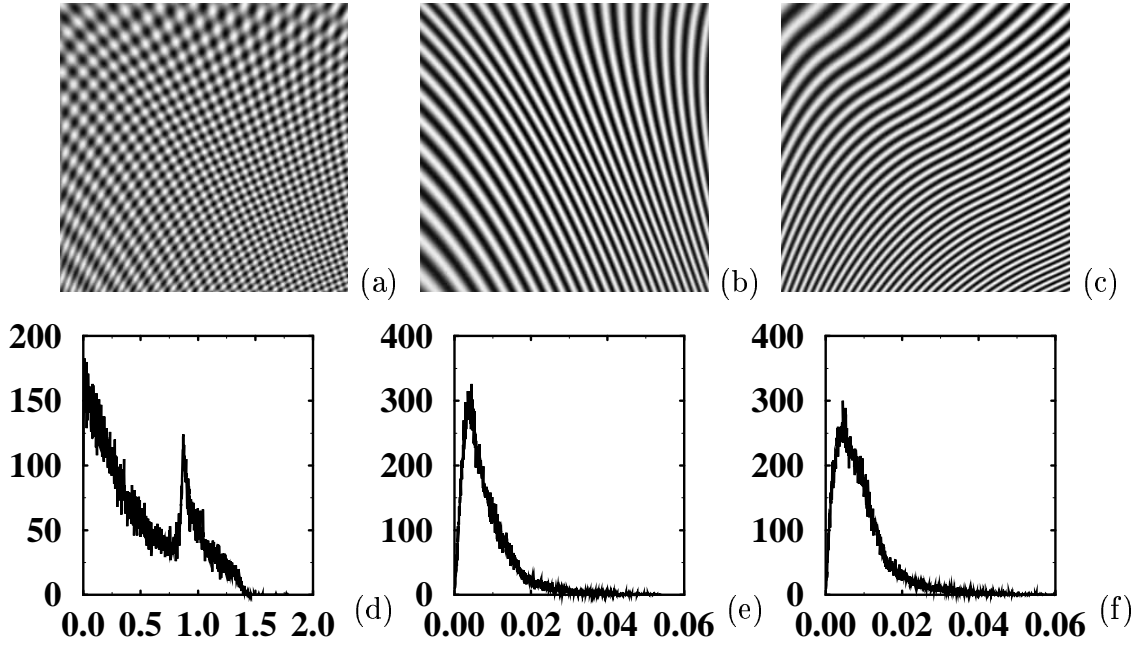


Figure 3: Nonstationary multi-component synthetic image showing one possible decomposition into highly locally coherent components. (a) Composite image. (b) Component zero. (c) Component one. (d) Histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for the composite image. (e) Histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for component zero. (f) Histogram of $B(\mathbf{x})^{\frac{1}{2}}$ for component one. Note the difference in the abscissa scale between the (d) part of the figure and the (e) and (f) parts, which clearly indicates the multi-partite nature of this image.

components is shown in the (b) and (c) parts of the figure. Figure 3 (d) shows the histogram of $\sqrt{B(\mathbf{x})}$ for the composite image, while histograms of $\sqrt{B(\mathbf{x})}$ for the individual components (shown in the (b) and (c) parts of the figure) are given in parts (d) and (e), respectively. Note that the abscissa scale for the bandwidth histogram of the composite image is two orders of magnitude larger than those for the bandwidth histograms of the individual components. Therefore, this is clearly a multi-component image. However, these observations suggest neither an approach for determining the number of components present in a general image, nor for performing the decomposition into components, both of which are important problems inviting future work.

We turn our attention now to simultaneous estimation of the modulating functions of *all* components of a multi-component image. The image is processed with a bank of filters of the form (11). By design, the filters are spaced densely in frequency so that virtually every

point in the frequency plane is covered by a filter responding at η -peak or higher. For a given image component, this ensures that the instantaneous frequency of the component almost always lies within the η -peak bandwidth of at least one channel filter. The filters must also be spatio-spectrally localized, so that the instantaneous frequencies of all other components lie far enough outside the η -peak bandwidth that the channel response is dominated by the given component. The spatially localized filtered demodulation algorithm (13), (14) is applied to the responses of *all* channels in the filterbank. The channel with response dominated by the given component then yields estimates of that component's modulating functions. Provided that each channel is dominated by at most one component at each pixel, estimates of the modulating functions of *all* components are produced at *all* pixels with this scheme. In image regions where all components are widely spaced in frequency, good estimates of the modulating functions of each component may even be produced by multiple channels.

At each pixel, there may, in general, also be channels which contain no image component within their η -peak bandwidth, but which contain multiple components *outside* their η -peak bandwidth such that no single component dominates the channel response. As a consequence of the nonlinearity of the filtered demodulation algorithm, cross-component interference will generally render the amplitude and frequency modulation estimates produced by such channels unusable for computation of multi-component AM-FM representations.

3 Computation of the Representation

To compute the multi-component AM-FM representation of an image, the image is processed with a multiband filter bank such as the one shown in Figure 2, and the filtered demodulation algorithm (13), (14) is applied to the response of each channel filter at every pixel. Hence, every channel in the filter bank produces observations $\hat{a}(\mathbf{x})$ and $\nabla \hat{\varphi}(\mathbf{x})$ at every pixel in the image. The problem of computing the representation then becomes one of determining the number of components present at each pixel and determining which channel should be used to estimate the modulating functions of *each* component at *each* pixel. Figure 4 depicts, in the frequency domain, a case where two components exist along a path in the image comprising $N + 1$ pixels. We impose a 1D ordering on the pixels, such that traversing them in order

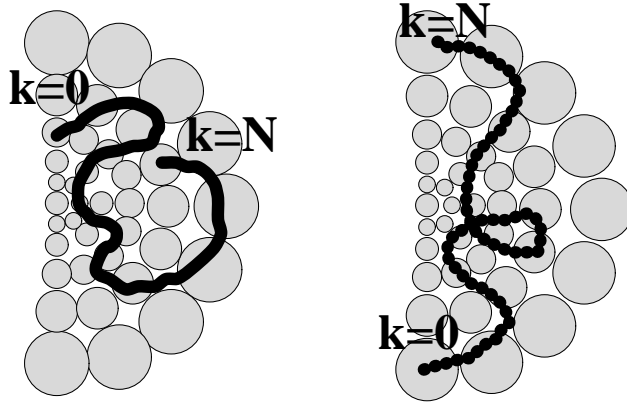


Figure 4: Frequency tracks of two image components in a region. The index k advances as we traverse the image pixels in the spatial domain.

results in our following a contiguous path in the spatial domain. Suppose we index the steps of the traversal with the variable k . Then, for each value of k , the instantaneous frequency $\nabla\varphi(\mathbf{x})$ of each component defines a point in the frequency plane. We can visualize this point tracing out a path as k grows, and this path may run across the frequency responses of the various filterbank channels. Indeed, as we traverse the pixels in order, k varies from 0 to N and, for each component, $\nabla\varphi(\mathbf{x})$ maps out a path in the frequency domain as depicted in Figure 4. In this section, we shall model the modulating functions of a component as affine functions of the state vector of a finite-dimensional linear system driven by uncorrelated noise. The estimates of the modulating functions produced by the algorithm (13), (14) will be modeled as noisy observations of an affine functions of the state vector, and a Kalman filter will be designed to yield optimal estimates of the modulating functions given the noisy observations as inputs.

3.1 Statistical State-Space Component Model

In this section we develop a statistical state-space model for an image component; this model will be used to assimilate the estimates $\hat{a}(\mathbf{x})$ and $\nabla\hat{\varphi}(\mathbf{x})$ produced from the filterbank channel responses into estimates of the modulating functions for each component. Using (1), the model for an image component is completely specified by the estimates of the amplitude modulation $a(\mathbf{x})$ and the estimates of the horizontal and vertical components of the frequency

modulation $\varphi^x(\mathbf{x}) = \frac{\partial}{\partial x}\varphi(\mathbf{x})$ and $\varphi^y(\mathbf{x}) = \frac{\partial}{\partial y}\varphi(\mathbf{x})$, up to an arbitrary constant of integration which may be specified by an initial value of the instantaneous phase at a single pixel. In developing the state-space model, we introduce an artificial temporal causality relationship between points \mathbf{x} in the sampled domain of the image by *ordering* them along a contiguous path of pixels, as discussed in conjunction with Figure 4. Specifically, the temporal causality relationship is induced by mapping points in the sampled spatial domain to a 1D lattice according to a path function $\mathcal{O} : \mathbf{x} \mapsto k$, where $k \in \mathbb{N}$ (we include the singleton $\{0\}$ in the naturals denoted by \mathbb{N}). This reparamaterization of the domain maps the three modulating functions of a component onto the 1D lattice according to:

$$a(\mathbf{x}) \xrightarrow{\mathcal{O}} a(k), \quad (16)$$

$$\varphi^x(\mathbf{x}) \xrightarrow{\mathcal{O}} \varphi^x(k), \quad (17)$$

$$\varphi^y(\mathbf{x}) \xrightarrow{\mathcal{O}} \varphi^y(k). \quad (18)$$

Let ρ denote continuous-domain arc length along the 1D path induced by \mathcal{O} , and use the notation

$$a'(k) = \left. \frac{\partial}{\partial \rho} a(\rho) \right|_{\rho=k}, \quad (19)$$

$$\varphi^{x'}(k) = \left. \frac{\partial}{\partial \rho} \varphi^x(\rho) \right|_{\rho=k}, \quad (20)$$

$$\varphi^{y'}(k) = \left. \frac{\partial}{\partial \rho} \varphi^y(\rho) \right|_{\rho=k}, \quad (21)$$

to indicate the restriction to the discrete 1D lattice of the derivatives of the modulating functions taken with respect to arc length along the path. Then it is possible to expand each modulating function in a first-order Taylor series about a lattice point k [38]:

$$a(k+1) = a(k) + a'(k) + \int_k^{k+1} (k+1-\rho) \frac{\partial^2}{\partial \rho^2} a(\rho) d\rho, \quad (22)$$

$$\varphi^x(k+1) = \varphi^x(k) + \varphi^{x'}(k) + \int_k^{k+1} (k+1-\rho) \frac{\partial^2}{\partial \rho^2} \varphi^x(\rho) d\rho, \quad (23)$$

$$\varphi^y(k+1) = \varphi^y(k) + \varphi^{y'}(k) + \int_k^{k+1} (k+1-\rho) \frac{\partial^2}{\partial \rho^2} \varphi^y(\rho) d\rho. \quad (24)$$

Likewise, the first-order derivatives of the modulating functions may be expanded in zeroth-order Taylor series

$$a'(k+1) = a'(k) + \int_k^{k+1} \frac{\partial^2}{\partial \rho^2} a(\rho) d\rho, \quad (25)$$

$$\varphi^{x'}(k+1) = \varphi^{x'}(k) + \int_k^{k+1} \frac{\partial^2}{\partial \rho^2} \varphi^x(\rho) d\rho, \quad (26)$$

$$\varphi^{y'}(k+1) = \varphi^{y'}(k) + \int_k^{k+1} \frac{\partial^2}{\partial \rho^2} \varphi^y(\rho) d\rho. \quad (27)$$

Under the usual assumptions that the derivatives of the modulating functions are small, slowly varying, or band limited [5, 6, 10, 13–17, 19–22, 33], an argument could be made for *truncating* the Taylor series (22) - (27). This amounts to considering that the modulating functions are *approximately* described by a *constant velocity* model:

$$a(k+1) \approx a(k) + a'(k), \quad (28)$$

$$a'(k+1) \approx a'(k), \quad (29)$$

$$\varphi^x(k+1) \approx \varphi^x(k) + \varphi^{x'}(k), \quad (30)$$

$$\varphi^{x'}(k+1) \approx \varphi^{x'}(k), \quad (31)$$

$$\varphi^y(k+1) \approx \varphi^y(k) + \varphi^{y'}(k), \quad (32)$$

$$\varphi^{y'}(k+1) \approx \varphi^{y'}(k). \quad (33)$$

However, the series (22) - (27) are *explicit*, and we obtain a more realistic image model here by allowing the constant velocity terms in (28) - (33) to *drift* by stochastic processes which approximate the integral terms in the explicit Taylor series. Specifically, we consider that $a(\mathbf{x})$ and $\varphi(\mathbf{x})$ are homogeneous, m.s. differentiable random fields, and model the integrals in (22) - (24) by three noise processes $u_a(k)$, $u_{\varphi_x}(k)$, and $u_{\varphi_y}(k)$, respectively. Likewise, we model the integrals in (25) - (27) with noise processes $\nu_a(k)$, $\nu_{\varphi_x}(k)$, and $\nu_{\varphi_y}(k)$, respectively. Collectively, we refer to these six noise processes as the *modulation accelerations*, or *MA* 's, since they involve local averages of the second derivatives of the modulating functions. Then, (22) - (27) may be written together in a canonical state-variable form to obtain the

statistical state-space component model

$$\begin{bmatrix} a(k+1) \\ a'(k+1) \\ \varphi^x(k+1) \\ \varphi^{x'}(k+1) \\ \varphi^y(k+1) \\ \varphi^{y'}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a(k) \\ a'(k) \\ \varphi^x(k) \\ \varphi^{x'}(k) \\ \varphi^y(k) \\ \varphi^{y'}(k) \end{bmatrix} + \begin{bmatrix} u_a(k) \\ \nu_a(k) \\ u_{\varphi_x}(k) \\ \nu_{\varphi_x}(k) \\ u_{\varphi_y}(k) \\ \nu_{\varphi_y}(k) \end{bmatrix}, \quad (34)$$

with output vector

$$\mathbf{Y}(k) = \begin{bmatrix} a(k) \\ \varphi^x(k) \\ \varphi^y(k) \end{bmatrix}. \quad (35)$$

Next, we must relate the quantities in the state-space model output vector $\mathbf{Y}(k)$ to the observations $\hat{a}(k)$, $\nabla \hat{\varphi}(k)$ obtained from the filtered demodulation algorithm (13), (14). Three factors contribute to errors in $\hat{a}(k)$ and $\nabla \hat{\varphi}(k)$: 1) discrete effects arising from the discretization of the algorithm (13), (14), 2) cross-component interference occurring in image regions where filterbank channel responses are dominated by more than one component, and 3) errors in the quasi-eigenfunction approximation inherent in the numerator and denominator of (13). We refer to these errors collectively as *measurement errors*, and model them with three noise processes $n_a(k)$, $n_{\varphi_x}(k)$, and $n_{\varphi_y}(k)$, called the *measurement noises*. This results in the observation equation

$$\begin{bmatrix} \hat{a}(k) \\ \widehat{\varphi^x}(k) \\ \widehat{\varphi^y}(k) \end{bmatrix} = \begin{bmatrix} a(k) \\ \varphi^x(k) \\ \varphi^y(k) \end{bmatrix} + \begin{bmatrix} n_a(k) \\ n_{\varphi_x}(k) \\ n_{\varphi_y}(k) \end{bmatrix}. \quad (36)$$

Note that we have not yet addressed the question of how to determine which channel of the multiband filterbank should be used to obtain the estimates $\hat{a}(k)$, $\nabla \hat{\varphi}(k)$ corresponding to a particular image component at any point k along the path induced by \mathcal{O} . If the nine noise processes $u_{\square}(k)$, $\nu_{\square}(k)$, and $n_{\square}(k)$ are appropriately specified, then the state-space model (34) - (36) is statistically valid for *any* homogeneous, m.s differentiable image component, under the assumption that the filterbank effectively isolates components on a

spatio-spectrally localized basis.

3.2 Specializing the Model

We now specialize the state space model (34) - (36) by making assumptions on the involved stochastic processes to arrive at a practical technique for computing the multi-component AM-FM representations of images belonging to specific classes. The validity of each assumption will be discussed in Section 3.2.5.

3.2.1 First-order Moments of Modulation Accelerations

We assume that $a(\mathbf{x})$ is twice m.s. differentiable and that $\varphi(\mathbf{x})$ is three times m.s. differentiable. Then, homogeneity implies that $E[a''(\mathbf{x})] = E[\varphi^{x''}(\mathbf{x})] = E[\varphi^{y''}(\mathbf{x})] = 0$. It follows immediately that

$$E[u_a(k)] = E[u_{\varphi_x}(k)] = E[u_{\varphi_y}(k)] = E[\nu_a(k)] = E[\nu_{\varphi_x}(k)] = E[\nu_{\varphi_y}(k)] = 0. \quad (37)$$

3.2.2 Second-order Moments of Modulation Accelerations

We use $C_{\square}(\cdot)$ to denote covariance functions between various stochastic processes, *e.g.*

$$C_{u_a, \nu_a}(k, j) = E[u_a(k)\nu_a(j)] - E[u_a(k)]E[\nu_a(j)]. \quad (38)$$

We assume that $\varphi(\mathbf{x})$ possesses a mildly restraining statistical property known as *quadrant symmetry* [39]. This property means that the covariance function of $\varphi(\mathbf{x})$ is even-symmetric in *each* component of the lag vector. While it *does not* suggest that the covariance structure of the phase of a real image should be *isotropic*, it does require that the covariation between the phase at pixel (i, j) and the phase at pixel $(i + m, j + n)$ be the same as that between the phase at pixel (i, j) and the phase at each of pixel $(i - m, j + n)$, pixel $(i + m, j - n)$, and pixel $(i - m, j - n)$, in a statistical sense. Thus, the assumption of quadrant symmetric phase is reasonable for a model that is to describe large classes of *locally coherent* image components. Quadrant symmetry implies that *even-order* cross-partials of $C_{\varphi}(\mathbf{x})$ are zero at the origin, and hence guarantees that all same-order cross-partials of $\varphi(\mathbf{x})$ are uncorrelated

at the lag space origin.

We further assume that the covariance structures of $a''(\mathbf{x})$, $\varphi^{x''}(\mathbf{x})$, and $\varphi^{y''}(\mathbf{x})$ are highly spatially localized, so that they effectively have impulse-like behavior in both the vertical and horizontal directions when viewed at the scale of the spatial sampling lattice (and, consequently, also at the scale of the discrete 1D lattice induced by \mathcal{O}). This assumption asserts that, across broad classes of spatially sampled locally coherent image components, we do not expect to find statistical correlation between the second derivative of the amplitude from pixel to pixel, nor between the third derivatives of the phase from pixel to pixel. Thus, these covariance structures may be modeled as being *effectively* impulsive *at the scale of the spatial sampling lattice*. However, when viewed at a *much* finer scale in the continuous-domain image component (*i.e.* prior to spatial sampling), the covariance structures of the partials of $a(\mathbf{x})$ and $\varphi(\mathbf{x})$ must be differentiable to guarantee existence of $\sigma_{a''}^2$, $\sigma_{\varphi^{x''}}^2$, and $\sigma_{\varphi^{y''}}^2$, the variances of $a''(\mathbf{x})$, $\varphi^{x''}(\mathbf{x})$, and $\varphi^{y''}(\mathbf{x})$. Hence, we are modeling the covariance structures of the partials of $a(\mathbf{x})$ and $\varphi(\mathbf{x})$ as smooth, localized bumps. This is consistent with the fact that, for locally coherent image components, we expect the modulating functions to be smoothly varying and/or band limited, so that their partials (particularly their *higher-order* partials) may be characterized having relatively small magnitudes (as compared to the magnitudes of the modulating functions themselves) and effectively random second-order statistical structure.

Then, the cross-covariance between $u_a(k)$ and $u_a(j)$ may be computed as follows:

$$\begin{aligned}
C_{u_a}(k, j) &= E[u_a(k)u_a(j)] \\
&= \sigma_{a''}^2 \delta(k - j) \int_k^{k+1} (k + 1 - \beta) \int_k^{k+1} (k + 1 - \alpha) \delta(\beta - \alpha) d\alpha d\beta \\
&= \frac{1}{3} \sigma_{a''}^2 \delta(k - j).
\end{aligned} \tag{39}$$

Analogous computations establish that

$$C_{u_{\varphi_x}}(k, j) = E[u_{\varphi_x}(k)u_{\varphi_x}(j)] = \frac{1}{3} \sigma_{\varphi^{x''}}^2 \delta(k - j), \tag{40}$$

$$C_{u_{\varphi_y}}(k, j) = E[u_{\varphi_y}(k)u_{\varphi_y}(j)] = \frac{1}{3} \sigma_{\varphi^{y''}}^2 \delta(k - j). \tag{41}$$

The cross-covariance between $\nu_a(k)$ and $\nu_a(j)$ is

$$\begin{aligned}
C_{\nu_a}(k, j) &= E[\nu_a(k)\nu_a(j)] \\
&= \sigma_{a''}^2 \delta(k - j) \int_k^{k+1} \int_k^{k+1} \delta(\beta - \alpha) d\alpha d\beta \\
&= \sigma_{a''}^2 \delta(k - j),
\end{aligned} \tag{42}$$

and it follows from similar computations that

$$C_{\nu_{\varphi_x}}(k, j) = E[\nu_{\varphi_x}(k)\nu_{\varphi_x}(j)] = \sigma_{\varphi_x''}^2 \delta(k - j), \tag{43}$$

$$C_{\nu_{\varphi_y}}(k, j) = E[\nu_{\varphi_y}(k)\nu_{\varphi_y}(j)] = \sigma_{\varphi_y''}^2 \delta(k - j). \tag{44}$$

Three more similar integrations yield

$$C_{u_a \nu_a}(k, j) = E[u_a(k)\nu_a(j)] = \frac{1}{2} \sigma_{a''}^2 \delta(k - j), \tag{45}$$

$$C_{u_{\varphi_x} \nu_{\varphi_x}}(k, j) = E[u_{\varphi_x}(k)\nu_{\varphi_x}(j)] = \frac{1}{2} \sigma_{\varphi_x''}^2 \delta(k - j), \tag{46}$$

$$C_{u_{\varphi_y} \nu_{\varphi_y}}(k, j) = E[u_{\varphi_y}(k)\nu_{\varphi_y}(j)] = \frac{1}{2} \sigma_{\varphi_y''}^2 \delta(k - j). \tag{47}$$

The quadrant symmetric structure of $\varphi(\mathbf{x})$ implies that

$$\begin{aligned}
E[u_{\varphi_x}(k)u_{\varphi_y}(j)] &= E[u_{\varphi_x}(k)\nu_{\varphi_y}(j)] = E[\nu_{\varphi_x}(k)u_{\varphi_y}(j)] \\
&= E[\nu_{\varphi_x}(k)\nu_{\varphi_y}(j)] = 0.
\end{aligned} \tag{48}$$

Finally, we recall that the amplitude modulation $a(\mathbf{x})$ corresponds to slowly varying contrast changes across a locally coherent image component, whereas information describing the local texture structure is manifest in the frequency modulation $\nabla\varphi(\mathbf{x})$. Since in general there is no reason to expect statistical correlation between contrast and local texture structure in a locally coherent image component, we assume that the amplitude $a(\mathbf{x})$ and instantaneous frequencies $\nabla\varphi(\mathbf{x})$ are independent. This implies

$$E[u_a(k)u_{\varphi_x}(j)] = E[u_a(k)u_{\varphi_y}(j)] = E[u_a(k)\nu_{\varphi_x}(j)]$$

$$\begin{aligned}
&= E[u_a(k)\nu_{\varphi_y}(j)] = E[\nu_a(k)u_{\varphi_x}(j)] \\
&= E[\nu_a(k)u_{\varphi_y}(j)] = E[\nu_a(k)\nu_{\varphi_x}(j)] \\
&= E[\nu_a(k)\nu_{\varphi_y}(j)] = 0.
\end{aligned} \tag{49}$$

3.2.3 Measurement Noise

Characterizing the three measurement noises $n_a(k)$, $n_{\varphi_x}(k)$, and $n_{\varphi_y}(k)$ of eq. (36) is difficult. We continue to assume that the response of each channel in the multiband filterbank is dominated by at most one image component at each pixel and to ignore issues concerning the discretization of the filtered demodulation algorithm (13), (14). Then, the measurement noises represent estimation errors arising *only* from errors in the quasi-eigenfunction approximation inherent in the numerator and denominator of (13). The errors in the frequency estimates are

$$n_{\varphi_x}(\mathbf{x}) = \varphi^x(\mathbf{x}) - \frac{A(x; \mathbf{x})}{B(\mathbf{x}) + C(\mathbf{x})} \tag{50}$$

and

$$n_{\varphi_y}(\mathbf{x}) = \varphi^y(\mathbf{x}) - \frac{A(y; \mathbf{x})}{B(\mathbf{x}) + C(\mathbf{x})}, \tag{51}$$

while the amplitude measurement error is

$$n_a(\mathbf{x}) = a(\mathbf{x}) - \sqrt{B^2(\mathbf{x}) + C^2(\mathbf{x})} \left| G_m \left(\frac{1}{B(\mathbf{x}) + C(\mathbf{x})} [A(x; \mathbf{x}) \quad A(y; \mathbf{x})]^T \right) \right|^{-1}, \tag{52}$$

where

$$\begin{aligned}
A(x; \mathbf{x}) &= \int_{\mathbb{R}^2} a(\mathbf{x} - \mathbf{y}) \Phi^T(\mathbf{x} - \mathbf{y}) \mathcal{G}^\dagger(m; \mathbf{y}) d\mathbf{y} \\
&\times \left\{ \int_{\mathbb{R}^2} \Phi^T(\mathbf{x} - \mathbf{y}) \left[a^x(\mathbf{x} - \mathbf{y}) \mathcal{G}^\dagger(m; \mathbf{y}) - a(\mathbf{x} - \mathbf{y}) \varphi^x(\mathbf{x} - \mathbf{y}) \mathcal{G}_\dagger(m; \mathbf{y}) \right] d\mathbf{y} \right. \\
&\left. - \int_{\mathbb{R}^2} \Phi^T(\mathbf{x} - \mathbf{y}) \left[a^x(\mathbf{x} - \mathbf{y}) \mathcal{G}_\dagger(m; \mathbf{y}) + a(\mathbf{x} - \mathbf{y}) \varphi^x(\mathbf{x} - \mathbf{y}) \mathcal{G}^\dagger(m; \mathbf{y}) \right] d\mathbf{y} \right\} \tag{53}
\end{aligned}$$

$$B(\mathbf{x}) = \int_{\mathbb{R}^2} a(\mathbf{x} - \mathbf{y}) \Phi^T(\mathbf{x} - \mathbf{y}) \mathcal{G}_\dagger(m; \mathbf{y}) d\mathbf{y}, \tag{54}$$

$$C(\mathbf{x}) = \int_{\mathbb{R}^2} a(\mathbf{x} - \mathbf{y}) \Phi^T(\mathbf{x} - \mathbf{y}) \mathcal{G}^\dagger(m; \mathbf{y}) d\mathbf{y}, \tag{55}$$

$$\Phi(\mathbf{x}) = \begin{bmatrix} \cos[\varphi(\mathbf{x})] \\ \sin[\varphi(\mathbf{x})] \end{bmatrix}, \quad (56)$$

$$\mathcal{G}^{\natural}(m; \mathbf{x}) = \begin{bmatrix} \text{Re}[g_m(\mathbf{x})] \\ -\text{Im}[g_m(\mathbf{x})] \end{bmatrix}, \quad (57)$$

$$\mathcal{G}_{\natural}(m; \mathbf{x}) = \begin{bmatrix} \text{Im}[g_m(\mathbf{x})] \\ \text{Re}[g_m(\mathbf{x})] \end{bmatrix}, \quad (58)$$

the observations at domain point \mathbf{x} were taken from filterbank channel m , $g_m(\mathbf{x})$ is the impulse response of the m^{th} channel filter, and $G_m(\boldsymbol{\Omega})$ is the 2D Fourier transform of $g_m(\mathbf{x})$. The expressions for the measurement noises in (50) - (52) are extremely cumbersome, and their joint and marginal moments are difficult to evaluate. Evaluation of their joint moments with the modulation accelerations are is equally difficult.

For locally coherent image components, the magnitudes of the second derivatives of the modulating functions should be very small or negligible in relation to the magnitudes of the actual modulating functions themselves. Indeed, with relation to (22) - (27), the second derivatives of the modulating functions might be interpreted as “higher-order terms.” Furthermore, in dealing with locally coherent signals and images, we have found that the QEA errors are normally quite small in practice [35, 40], except possibly at a few isolated points corresponding to phase discontinuities in the image. The effects of QEA errors at these points can be ameliorated by smoothing the estimated instantaneous frequencies with a low-pass filter [4, 5]. Therefore, despite the fact that closed form solutions for the covariances between the measurement noises in (36) and the MA’s in (34) are not available, we expect that the magnitudes of any covariance that might exist should be very slight, especially if postfiltering of the estimated modulating functions is employed, and we assume for the remainder of this report that the MA’s and measurement noises are uncorrelated. In the event that later research establishes significant correlations between the MA’s and measurement noises, and these correlations can be quantified, then the model (34) - (36) will be improved by incorporating the measurement noises into the state equation (34). This would lead to an extended Kalman filter for the optimal estimator of the modulating functions.

It is extremely difficult to infer anything about correlations among the measurement noises from (50) - (52). However, given the block-diagonal structure of the state transition

matrix in (34) and the lack of correlation between the MA's, we do not expect significant covariance between n_{φ_x} and n_{φ_y} . In view of (14), it is apparent that errors in the frequency estimates will lead to errors in the amplitude estimates. However, given the lack of closed form expressions for these correlations, we will neglect them and realize that the model (34) - (36) is suboptimal in this sense.

3.2.4 State Vector Covariance Matrix

We use the notation $\boldsymbol{\pi}(k)$ to denote the state vector covariance matrix associated with (34):

$$\boldsymbol{\pi}(k) = E \begin{bmatrix} a(k) \\ a'(k) \\ \varphi^x(k) \\ \varphi^{x'}(k) \\ \varphi^y(k) \\ \varphi^{y'}(k) \end{bmatrix} \begin{bmatrix} a(k) \\ a'(k) \\ \varphi^x(k) \\ \varphi^{x'}(k) \\ \varphi^y(k) \\ \varphi^{y'}(k) \end{bmatrix}^T. \quad (59)$$

The (initial) value of the state vector covariance matrix at the first pixel along the path induced by \mathcal{O} where a component is supported is denoted by $\boldsymbol{\pi}(0)$.

3.2.5 Summary of Assumptions

In this section, we summarize the assumptions that have been made in Sections 3.2.1 - 3.2.3. We assumed that $a(\mathbf{x})$ is a homogeneous, twice m.s. differentiable random field and that $\varphi(\mathbf{x})$ is a three-times m.s. differentiable random field. The differentiability assumptions are reasonable, since any image may be considered as samples of a continuous-domain image with continuously differentiable modulating functions. The homogeneity assumptions are also reasonable in view of the law of large numbers. While the spatially computed sample statistics of the modulating functions of a specific image may vary with space, we wish the model (34) - (36) to be generally applicable to broad classes of images; it must be recalled that any *single* image is but a single realization of the stochastic model.

We also assumed in Section 3.2.2 that the correlation structure of $\varphi(\mathbf{x})$ is quadrant symmetric [39]. This means that the covariance function $C_\varphi(\mathbf{x})$ is even in each component

of the lag vector, and that the associated cross-spectral density functions are real-valued. The assumption says that, across a broad class of locally coherent image components, we expect that the covariation between the phase at one pixel and the phase at a second pixel is *statistically* the same as that between the phase at the first pixel and the phase at a third pixel separated from the first by an equal and opposite displacement. The assumption of quadrant symmetry is a mild one in view of the fact that we want the model (34) - (36) to be generally applicable across broad classes of images. Furthermore, quadrant symmetry arises naturally in many applications. For example, images of scenes arising from physical processes governed by isotropic, separable, or ellipsoidal random fields are all expected to admit quadrant symmetric correlation structures.

We assumed that $a(\mathbf{x})$ and $\nabla\varphi(\mathbf{x})$ are independent, which is justified since there is, in general, no reason to expect the contrast and local frequency content of an image to depend on one another. Consistent with the usual assumptions that the modulating functions have small derivatives [5, 6, 14, 17, 33], or are band limited [10, 13, 15, 16, 19–22], we additionally argued that we expected the partials of the modulating functions to have small magnitudes with relation to the magnitudes of the modulating functions themselves, and that we also expected the partials to be random in nature. It should be recalled that this *does not* imply that we expect the *sample* covariances computed from any single image component to be everywhere identically zero; rather, it implies that we expect a lack of spatial correlation between the modulating function partials computed over *broad classes* of locally coherent image components. Hence, we assumed that the correlation structures of $a''(\mathbf{x})$, $\varphi^{x''}(\mathbf{x})$, and $\varphi^{y''}(\mathbf{x})$ are spatially localized and possess impulse-like behavior at the scale of the spatial sampling lattice.

Finally, in Section 3.2.3 we assumed that the measurement noises of (36) are mutually uncorrelated and uncorrelated with the MA's. While closed form solutions for these covariances are not yet available, we argued informally that they should be small in relation to the magnitudes of the modulating functions.

Under the assumptions summarized in this section, the state-space model (34), (35) is no longer completely general. Indeed, once specific values are assigned to the MA noise powers, the measurement noise powers, and $\pi(0)$, the model is statistically valid only for classes

of image components whose variation conforms to the constraints of the model. However, irrespective of the image under consideration, the discrete sequences $a(k)$ and $\nabla\varphi(k)$ may be considered samples of twice continuously differentiable bounded functions $a(\rho)$, $\nabla\varphi(\rho)$. Therefore, existence and boundedness of all quantities in (22) - (27) can be guaranteed for *any* bounded discrete-domain image. Furthermore, for locally coherent image components, the variations in $a(k)$ and $\nabla\varphi(k)$ are necessarily smooth and not too erratic. Furthermore, the magnitudes of the MA's should be small compared to those of the modulating functions themselves. Therefore, we expect the model (34), (35) to work well for describing broad classes of locally coherent image components.

By example, Figure 5 shows the real parts of sixteen 256×256 8-bit grey scale image components generated from (34), (35), all using the arbitrarily selected values $\sigma_a'' = 5 \times 10^{-8}$, $\sigma_{\varphi x''} = \sigma_{\varphi y''} = 10^{-7}$, and

$$\boldsymbol{\pi}(0) = \text{diag} \{ 10^{-7} \quad 5 \times 10^{-8} \quad 5 \times 10^{-2} \quad 10^{-7} \quad 5 \times 10^{-2} \quad 10^{-7} \}. \quad (60)$$

The instantaneous phase of the first pixel (upper left corner) of each component was taken to be uniform in $[-\pi, \pi]$, and the path \mathcal{O} was specified by

$$k = \begin{cases} nM + m, & n \text{ even}; \\ (n+1)M - m - 1, & n \text{ odd}, \end{cases} \quad (61)$$

where M is the column dimension of the image and $\mathbf{x} = [m \quad n]^T$. This path corresponds to traversing back and forth along alternating image rows, as depicted in figure 6.

From Figure 5, it is clear that the statistical state-space AM-FM image component model is inherently capable of generating significant nonstationary features and structure. Also evident from the figure is the fact that wide ranges of spatial frequencies, and therefore broad classes of textured regions, can be characterized by the model, even for a single choice of the parameters.

We now address the problem of determining the measurement noise powers $\sigma_{n_a}^2$, $\sigma_{n_{\varphi_x}}^2$, and $\sigma_{n_{\varphi_y}}^2$ defined through (36). In [1], we described an analysis paradigm called *dominant component analysis*, which, for an image which *may* be multi-partite, estimates on a pointwise

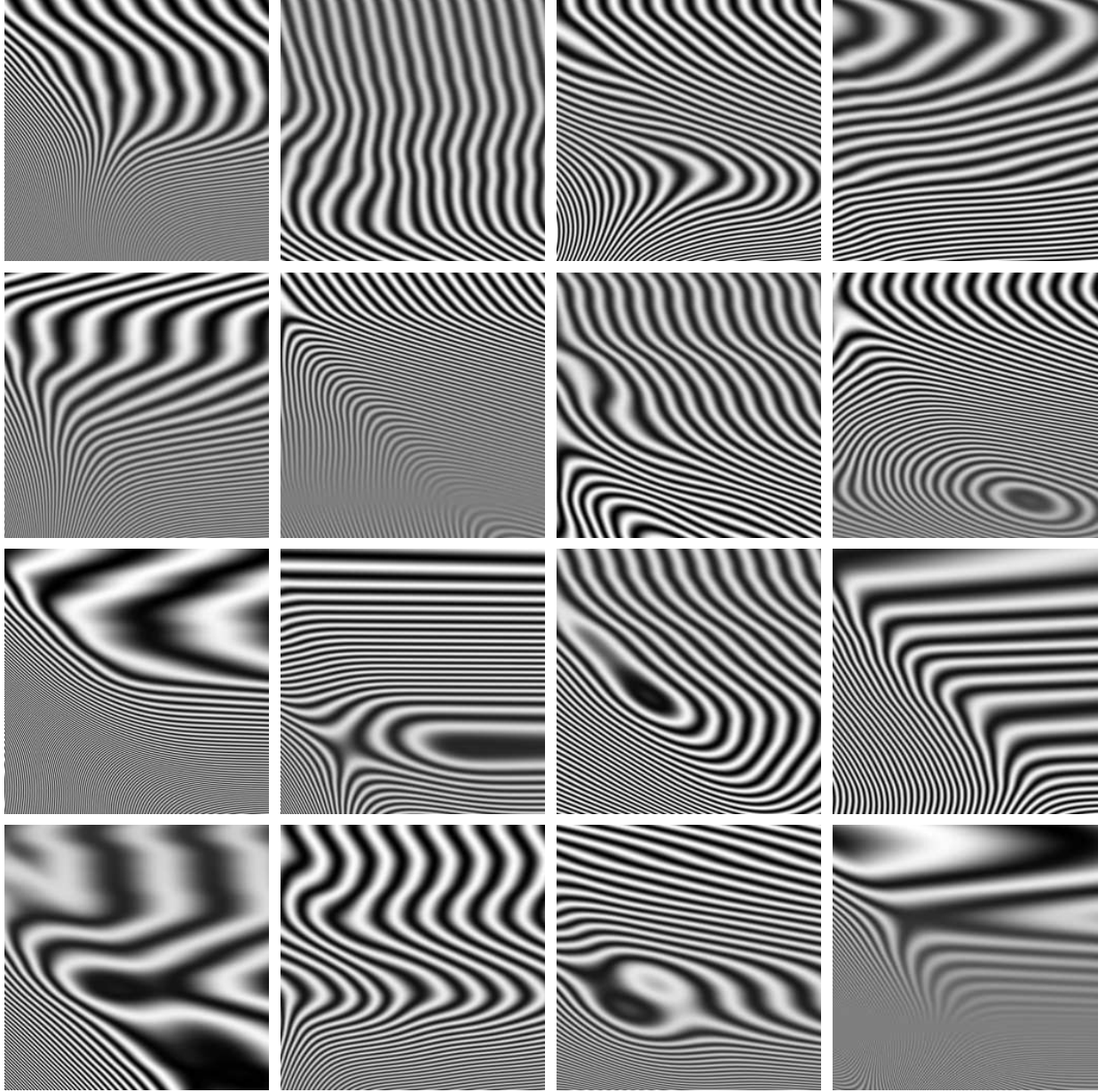


Figure 5: Real parts of 16 image components generated from the statistical state-space model using values $\sigma_{u_a} = 5 \times 10^{-8}$ and $\sigma_{u_{\varphi_x}} = \sigma_{u_{\varphi_y}} = 10^{-7}$ (see text for values of $\boldsymbol{\pi}(0)$).

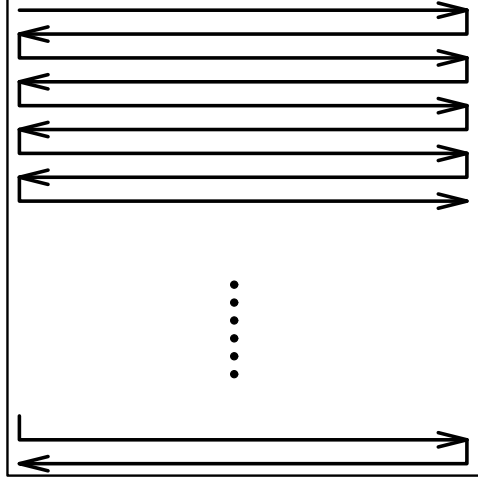


Figure 6: Path function $\mathcal{O} : \mathbf{x} \mapsto k$ used to generate the sixteen image components shown in Figure 5, as given by Equation (61).

basis the modulating functions corresponding to the component which dominates the local instantaneous frequency spectrum of the image. When this paradigm is applied to a *single-component* image, estimates $\hat{a}(k)$, $\nabla \hat{\varphi}(k)$ corresponding to those in (36) are obtained for the component at every pixel. Provided that the true values $a(k)$, $\nabla \varphi(k)$ are known and assuming that the image is ergodic, high quality estimates of the measurement noise powers can be obtained by subtracting the true values from the dominant component estimates and by computing the sample variance of the differences. Applying this approach to all 16 images shown in Figure 5, we obtained the values

$$\sigma_{n_a}^2 \approx 1.7379 \times 10^{-2}, \quad (62)$$

$$\sigma_{n_{\varphi_x}}^2 \approx 9.6432 \times 10^{-5}, \quad (63)$$

$$\sigma_{n_{\varphi_y}}^2 \approx 3.4027 \times 10^{-3}. \quad (64)$$

If the modulation acceleration noise powers, measurement noise powers, and initial state vector covariance matrix $\boldsymbol{\pi}(0)$ are known, then the problem of determining from which channel of the filterbank to take observations $\hat{a}(k)$ and $\nabla \hat{\varphi}(k)$ for any given image component can be solved by following that component's frequency domain track, as depicted in Figure 4, with a track processor. At each pixel k along the path induced by \mathcal{O} , the track processor

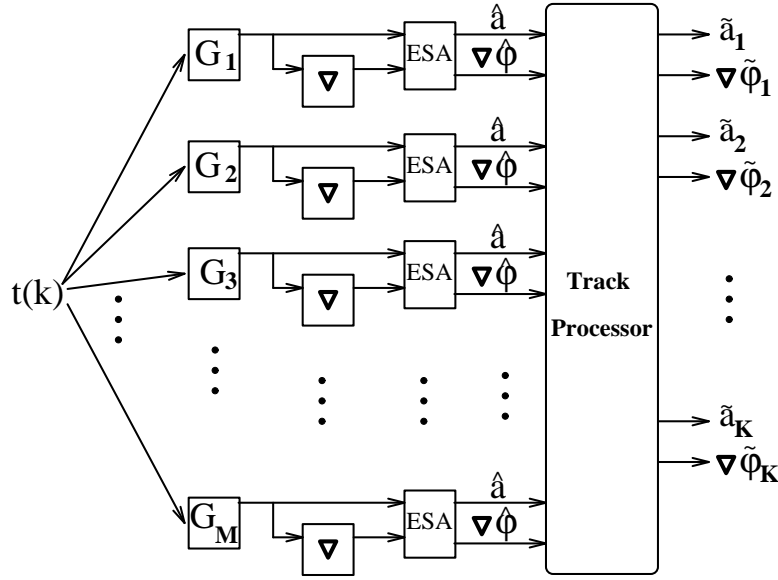


Figure 7: Block diagram of the tracked multi-component paradigm.

predicts $\nabla \hat{\varphi}(k)$ and $\hat{a}(k)$, and observations of these quantities are taken from the channel with center frequency closest to the predicted instantaneous frequencies. Optimal estimates of the modulating functions can then be computed based on the observations and the track history. A block diagram of the approach is shown in Figure 7. In the figure, the discrete-domain image $t(\mathbf{x})$ is analyzed with an M -channel filterbank, and estimates $\hat{a}(k)$, $\nabla \hat{\varphi}(k)$ are produced by each of the M channels at every pixel. From these observations, the track processor formulates optimal estimates $\tilde{a}(k)$, $\nabla \tilde{\varphi}(k)$ at every pixel for each of the K image components.

With the state-space model (34),(35), subject to the assumptions summarized above in this section, and observation equation (36), we have modeled the estimates of the modulating functions of an AM-FM image component produced by the filtered demodulation algorithm (13), (14) as noisy observations of an affine function of the state vector of a finite-dimensional linear system driven by uncorrelated noise. The Kalman filter is the unique linear system which, given such noisy observations, produces minimum mean squared error unbiased estimates of any affine function of the state vector of such a system. Therefore, we use a Kalman filter for the track processor in Figure 7. The filter design is described in Section 3.3.

3.3 Track Processor

In this section we develop the track processor shown in Figure 7. We begin by observing that the system modes of the model (34), (35) corresponding to the amplitude modulation, the horizontal component of the frequency modulation, and the vertical component of the frequency modulation can be decoupled to yield three independent second-order systems. Hence, an independent Kalman filter will be used to track each one. The system for the amplitude modulation is

$$\begin{bmatrix} a(k+1) \\ a'(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a(k) \\ a'(k) \end{bmatrix} + \begin{bmatrix} u_a(k) \\ \nu_a(k) \end{bmatrix}, \quad (65)$$

$$a(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} a(k) \\ u_a(k) \end{bmatrix}, \quad (66)$$

with observation equation

$$\hat{a}(k) = a(k) + n_a(k) \quad (67)$$

and initial state vector covariance matrix

$$\boldsymbol{\pi}_a(0) = E \begin{bmatrix} a(0) \\ a'(0) \end{bmatrix} \begin{bmatrix} a(0) \\ a'(0) \end{bmatrix}^T. \quad (68)$$

Likewise, the decoupled instantaneous frequency system models are

$$\begin{bmatrix} \varphi^x(k+1) \\ \varphi^{x'}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varphi^x(k) \\ \varphi^{x'}(k) \end{bmatrix} + \begin{bmatrix} u_{\varphi_x}(k) \\ \nu_{\varphi_x}(k) \end{bmatrix}, \quad (69)$$

$$\varphi^x(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \varphi^x(k) \\ \varphi^{x'}(k) \end{bmatrix}, \quad (70)$$

$$\widehat{\varphi^x}(k) = \varphi^x(k) + n_{\varphi_x}(k) \quad (71)$$

$$\boldsymbol{\pi}_{\varphi_x}(0) = E \begin{bmatrix} \varphi^x(0) \\ \varphi^{x'}(0) \end{bmatrix} \begin{bmatrix} \varphi^x(0) \\ \varphi^{x'}(0) \end{bmatrix}^T \quad (72)$$

and

$$\begin{bmatrix} \varphi^y(k+1) \\ \varphi^{y'}(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varphi^y(k) \\ \varphi^{y'}(k) \end{bmatrix} + \begin{bmatrix} u_{\varphi_y}(k) \\ \nu_{\varphi_y}(k) \end{bmatrix}, \quad (73)$$

$$\varphi^y(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \varphi^y(k) \\ \varphi^{y'}(k) \end{bmatrix}, \quad (74)$$

$$\widehat{\varphi}^y(k) = \varphi^y(k) + n_{\varphi_y}(k) \quad (75)$$

$$\boldsymbol{\pi}_{\varphi_y}(0) = E \begin{bmatrix} \varphi^y(0) \\ \varphi^{y'}(0) \end{bmatrix} \begin{bmatrix} \varphi^y(0) \\ \varphi^{y'}(0) \end{bmatrix}^T. \quad (76)$$

We now outline the design of the Kalman filter for the amplitude estimator. The MA covariance matrix is

$$E \begin{bmatrix} u_a(k) \\ \nu_a(k) \end{bmatrix} \begin{bmatrix} u_a(j) \\ \nu_a(j) \end{bmatrix}^T = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \sigma_{a''}^2 \delta(k-j), \quad (77)$$

while the measurement noise covariance is

$$E[n_a(k) \ n_a(j)] = \sigma_{n_a}^2 \delta(k-j). \quad (78)$$

The predicted state vector at pixel k along the path induced by \mathcal{O} given $k-1$ observations $\widehat{a}(j)$ is

$$\widetilde{X}(k|k-1) = \begin{bmatrix} \widetilde{a}(k|k-1) \\ \widetilde{a}'(k|k-1) \end{bmatrix}, \quad (79)$$

and the recursive formulation for the predicted state vector error covariance matrix is

$$P(k+1|k) = \begin{bmatrix} P_{1,1}(k+1|k) & P_{1,2}(k+1|k) \\ P_{2,1}(k+1|k) & P_{2,2}(k+1|k) \end{bmatrix} \quad (80)$$

where

$$\begin{aligned} P_{1,1}(k+1|k) &= P_{1,1}(k|k-1) + 2P_{1,2}(k|k-1) + P_{2,2}(k|k-1) \\ &\quad - \frac{[P_{1,1}(k|k-1) + P_{1,2}(k|k-1)]^2}{P_{1,1}(k|k-1) + \sigma_{n_a}^2} + \frac{1}{3}\sigma_{a''}^2, \\ P_{1,2}(k+1|k) &= P_{2,1}(k+1|k) = P_{1,2}(k|k-1) + P_{2,2}(k|k-1) \end{aligned}$$

$$- \frac{P_{1,2}(k|k-1)[P_{1,1}(k|k-1) + P_{1,2}(k|k-1)]}{P_{1,1}(k|k-1) + \sigma_{na}^2} + \frac{1}{2}\sigma_{a''}^2,$$

and

$$P_{2,2}(k+1|k) = P_{2,2}(k|k-1) - \frac{P_{1,2}^2(k|k-1)}{P_{1,1}(k|k-1) + \sigma_{na}^2} + \sigma_{a''}^2.$$

The estimated state vector at pixel k given k observations is

$$\widetilde{X}(k|k) = \begin{bmatrix} \widetilde{a}(k|k) \\ \widetilde{a}'(k|k) \end{bmatrix}, \quad (81)$$

and its error covariance matrix is

$$P(k|k) = \begin{bmatrix} P_{1,1}(k|k-1) - \frac{P_{1,1}^2(k|k-1)}{P_{1,1}(k|k-1) + \sigma_{na}^2} & P_{1,2}(k|k-1) - \frac{P_{1,1}(k|k-1)P_{1,2}(k|k-1)}{P_{1,1}(k|k-1) + \sigma_{na}^2} \\ P_{1,2}(k|k-1) - \frac{P_{1,1}(k|k-1)P_{1,2}(k|k-1)}{P_{1,1}(k|k-1) + \sigma_{na}^2} & P_{2,2}(k|k-1) - \frac{P_{1,2}^2(k|k-1)}{P_{1,1}(k|k-1) + \sigma_{na}^2} \end{bmatrix}. \quad (82)$$

Finally, the Kalman gain vector for the amplitude estimator is

$$K(k) = \begin{bmatrix} \alpha_a(k) \\ \beta_a(k) \end{bmatrix} = \begin{bmatrix} \frac{P_{1,1}(k|k-1)}{P_{1,1}(k|k-1) + \sigma_{na}^2} \\ \frac{P_{2,1}(k|k-1)}{P_{1,1}(k|k-1) + \sigma_{na}^2} \end{bmatrix}. \quad (83)$$

Completely analogous developments lead to the gain sequences for the instantaneous frequency estimators, from which it follows that the optimal equations for the track processor in Figure 7 are

$$\widetilde{a}(k|k) = \widetilde{a}(k|k-1) + \alpha_a(k)(\widehat{a}(k) - \widetilde{a}(k|k-1)), \quad (84)$$

$$\widetilde{a}'(k+1|k) = \widetilde{a}'(k|k-1) + \beta_a(k)(\widehat{a}(k) - \widetilde{a}(k|k-1)), \quad (85)$$

$$\widetilde{a}(k+1|k) = \widetilde{a}(k|k) + \widetilde{a}'(k+1|k), \quad (86)$$

$$\widetilde{\varphi}^x(k|k) = \widetilde{\varphi}^x(k|k-1) + \alpha_{\varphi_x}(k)(\widehat{\varphi}^x(k) - \widetilde{\varphi}^x(k|k-1)), \quad (87)$$

$$\widetilde{\varphi}^{xt}(k+1|k) = \widetilde{\varphi}^{xt}(k|k-1) + \beta_{\varphi_x}(k)(\widehat{\varphi}^x(k) - \widetilde{\varphi}^x(k|k-1)), \quad (88)$$

$$\widetilde{\varphi}^x(k+1|k) = \widetilde{\varphi}^x(k|k) + \widetilde{\varphi}^{xt}(k+1|k), \quad (89)$$

$$\widetilde{\varphi}^y(k|k) = \widetilde{\varphi}^y(k|k-1) + \alpha_{\varphi_y}(k)(\widehat{\varphi}^y(k) - \widetilde{\varphi}^y(k|k-1)), \quad (90)$$

$$\widetilde{\varphi}^{yt}(k+1|k) = \widetilde{\varphi}^{yt}(k|k-1) + \beta_{\varphi_y}(k)(\widehat{\varphi}^y(k) - \widetilde{\varphi}^y(k|k-1)), \quad (91)$$

$$\widetilde{\varphi}^y(k+1|k) = \widetilde{\varphi}^y(k|k) + \widetilde{\varphi}^{yt}(k+1|k). \quad (92)$$

Subject to the assumptions outlined in Section 3.2.5, the track processor (84) - (92) is unbiased:

$$E \begin{bmatrix} \tilde{a}(k) \\ \widetilde{\varphi^x}(k) \\ \widetilde{\varphi^y}(k) \end{bmatrix} = \begin{bmatrix} a(k) \\ \varphi^x(k) \\ \varphi^y(k) \end{bmatrix}. \quad (93)$$

It is also optimal, in the sense that the mean squared errors in $\tilde{a}(k)$ and $\nabla \tilde{\varphi}(k)$ are minimized.

3.4 New Track Starts

At each point i in the path \mathcal{O} , we start new tracks using the observations $\hat{a}(i)$ and $\nabla \hat{\varphi}(i)$ obtained from channels which maximize the quantity

$$Q(n) = \frac{\hat{a}_n(i) |G_n(\nabla \hat{\varphi}_n(i))|}{\max_{\Omega} |G_n(\Omega)|}, \quad (94)$$

where $\hat{a}_n(i)$ and $\nabla \hat{\varphi}_n(i)$ are the amplitude and frequency estimates obtained from filterbank channel n , provided that these observations 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 channel whose center frequency is closest to the instantaneous frequency of the component, affording improved immunity against out-of-band information through an enhanced SNR.

The track is initialized at pixel i using channel n by taking

$$\tilde{a}(0|0) = \hat{a}_n(i) \quad (95)$$

$$\tilde{a}'(0|0) = \hat{a}_n(i) - \hat{a}_n(i-1) \quad (96)$$

$$\widetilde{\varphi^x}(0|0) = \widehat{\varphi^x}(i) \quad (97)$$

$$\widetilde{\varphi^{x'}}(0|0) = \widehat{\varphi^x}(i) - \widehat{\varphi^x}(i-1) \quad (98)$$

$$\widetilde{\varphi^y}(0|0) = \widehat{\varphi^y}(i) \quad (99)$$

$$\widetilde{\varphi^{y'}}(0|0) = \widehat{\varphi^y}(i) - \widehat{\varphi^y}(i-1) \quad (100)$$

Typically, for some small number L , new tracks are started at each pixel using the amplitude

and frequency observations from the L channels having the largest values of $Q(n)$ from among those channels producing observations that do not associate with an already existing track. Under the assumption that the filter bank has been constructed such 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 $\lceil K/L \rceil^{\text{th}}$ pixel of the region.

4 Reconstruction

In the preceding section, we described a practical technique for computing the multi-component AM-FM representation

$$\{\tilde{a}_k(m, n), \nabla \tilde{\varphi}_k(m, n)\}_{k \in [1, K], (m, n) \in [0, M-1] \times [0, N-1]} \quad (101)$$

of an M column by N row K -component discrete-domain image. Reconstruction of the image from the representation is a difficult, ill-posed inverse problem. If error free continuous-domain frequency estimates $\nabla \tilde{\varphi}_k(\mathbf{x})$ were available, then the instantaneous phase $\varphi_k(\mathbf{x})$ of the k^{th} component could be recovered by integrating along arbitrary paths in the image domain. However, due to estimation errors in (13) and in (84) - (92), the estimated phase gradient field is not, in practice, conservative. Therefore, summing the discrete-domain 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 could be employed, we have found that, for components supported in a rectangular region, it is often satisfactory to reconstruct the phase of the k^{th} component according to

$$\tilde{\varphi}_k(m, n) = \frac{1}{2} \left[\tilde{\varphi}_k(m-1, n) + \widetilde{\varphi}_k^x(m-1, n) + \tilde{\varphi}_k(m, n-1) + \widetilde{\varphi}_k^y(m, n-1) \right]. \quad (102)$$

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

$$\tilde{\varphi}_k(m, n_0) = \tilde{\varphi}_k(m-1, n_0) + \tilde{\varphi}_k^x(m-1, n_0) \quad (103)$$

and down the leftmost column by iterating

$$\tilde{\varphi}_k(m_0, n) = \tilde{\varphi}_k(m_0, n-1) + \tilde{\varphi}_k^y(m_0, n-1). \quad (104)$$

The phase reconstruction is initialized by saving an initial estimated phase sample

$$\tilde{\varphi}_k(m_0, n_0) = \arctan \left\{ \frac{\text{Im} [t_i(m_0, n_0)]}{\text{Re} [t_i(m_0, n_0)]} \right\} \quad (105)$$

during computation of the multi-component representation (101), where tracking of the k^{th} 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 i^{th} channel filter (the filters given by (10), (11) are conjugate symmetric in the space domain, and hence introduce no phase shift in the response image). Note that in (105) it is again necessary to assume that $t_i(m_0, n_0)$, the response of the channel used to start the track, is dominated by only one image component. Interference from other components would cause the phase reconstruction to be initialized on an erroneous value.

With the phase reconstruction algorithm (102) - (105), the deleterious effects of estimation errors in $\nabla \tilde{\varphi}_k(\mathbf{x})$ are cumulative. To ameliorate this effect, we improve the robustness of the technique by saving estimated phase samples $\tilde{\varphi}_k(m, n)$ on a rectangular grid. We then reconstruct the phase on each rectangle of the grid independently, beginning with the phase sample at the upper left corner. The component is then reconstructed using (1):

$$\tilde{t}_k(m, n) = \tilde{a}_k(m, n) \exp [j \tilde{\varphi}_k(m, n)]. \quad (106)$$

Once $\tilde{t}_k(m, n)$ is recovered for all components, the image is reconstructed from (9) according to

$$\tilde{t}(m, n) = \sum_{k=1}^K \tilde{t}_k(m, n). \quad (107)$$

5 Examples

The 256×256 two-component image of Figure 3 is shown again in Figure 8(a). Components one and two are shown individually in the (b) and (c) parts of Figure 8, respectively. The multi-component paradigm depicted in Figure 7 was applied to compute the multi-component AM-FM representation (101) using the multiband filterbank depicted in Figure 2, the filtered demodulation algorithm (13), (14), and the track processor described in Section 3.3. To avoid edge effects from the filterbank, estimation and component tracking were not performed on the outside 16 rows and columns of the image.

The track processor correctly identified the presence of two components. Amplitude estimates $\tilde{a}_1(m, n)$ for component one are shown in the (d) part of the figure, while the horizontal and vertical components of the instantaneous frequency estimates, $\widetilde{\varphi}_1^x(m, n)$, $\widetilde{\varphi}_1^y(m, n)$ are shown in the (e) and (f) parts, respectively. These estimated 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 rightmost portion of component one as depicted in Figure 8(b); the vertical frequency is very nearly zero in this region. The oscillations occur because, at very low frequencies, the distinction between which image features should be interpreted as amplitude modulation and which should be interpreted as frequency modulation becomes unclear. Note that this oscillatory behavior also propagates into the amplitude estimates, shown in Figure 8(d), as a consequence of the dependence of the amplitude estimates on the frequency estimates through Equation (14). The amplitude and frequency estimates $\tilde{a}_2(m, n)$, $\widetilde{\varphi}_2^x(m, n)$, and $\widetilde{\varphi}_2^y(m, n)$ for component two are shown in the (g), (h), and (i) parts of Figure 8. Once again, these estimates are in near perfect agreement with the true values. Collectively, the (d) – (i) parts of Figure 8 constitute the multi-component AM-FM representation (101) 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 from the multi-component AM-FM representation on a 32×32 pixel grid of estimated phase samples (105) using the algorithm (102) – (104), (106). The reconstructions of components one and two are shown in the (j) and (k)

parts of Figure 8, respectively. Visually, the reconstructed components are virtually indistinguishable from the true values. The reconstruction of the multi-component image using (107) is shown in the (l) part of the figure, and is also virtually indistinguishable from the true values in Figure 8(a).

Finally, we applied the multi-component paradigm to compute a multi-component AM-FM representation for the tree image of Figure 12 (a). This complicated, nonstationary multi-partite image presents significant challenges since the number of components, as well as the sizes and shapes of their regions of support are unknown. The filterbank of Figure 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 the demodulation algorithm (13), (14) due to cross-component interference. Furthermore, there is no way to determine the optimal Kalman gain sequences for the track processor, and hence the computed representation is suboptimal.

As we noted in the preceding example, 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 amplitude modulation, and which should be interpreted as frequency modulation. Hence, we began by extracting the low-frequency component shown in Figure 12(b) by linear filtering prior to computation of the representation. Next, we heuristically divided the image into rectangular regions, and applied the approach of Figure 7 to each one. Once again, component tracking and estimation were not performed on the outside 16 rows and columns of the image to avoid filterbank edge effects. Many tracks were generated in each region, and we hand-selected 41 of them.

Reconstructions of these 41 components computed on 4×4 pixel grids of estimated phase samples are shown in Figures 9 – 11, where each individual component image has been *independently* scaled for maximum dynamic range of grey scales. Summing these reconstructed components with the low-frequency component shown in Figure 12(b), we obtained the excellent reconstruction shown in Figure 12 (c). Clearly, the 41-component AM-FM representation has succeeded in capturing the essential structure, features, and

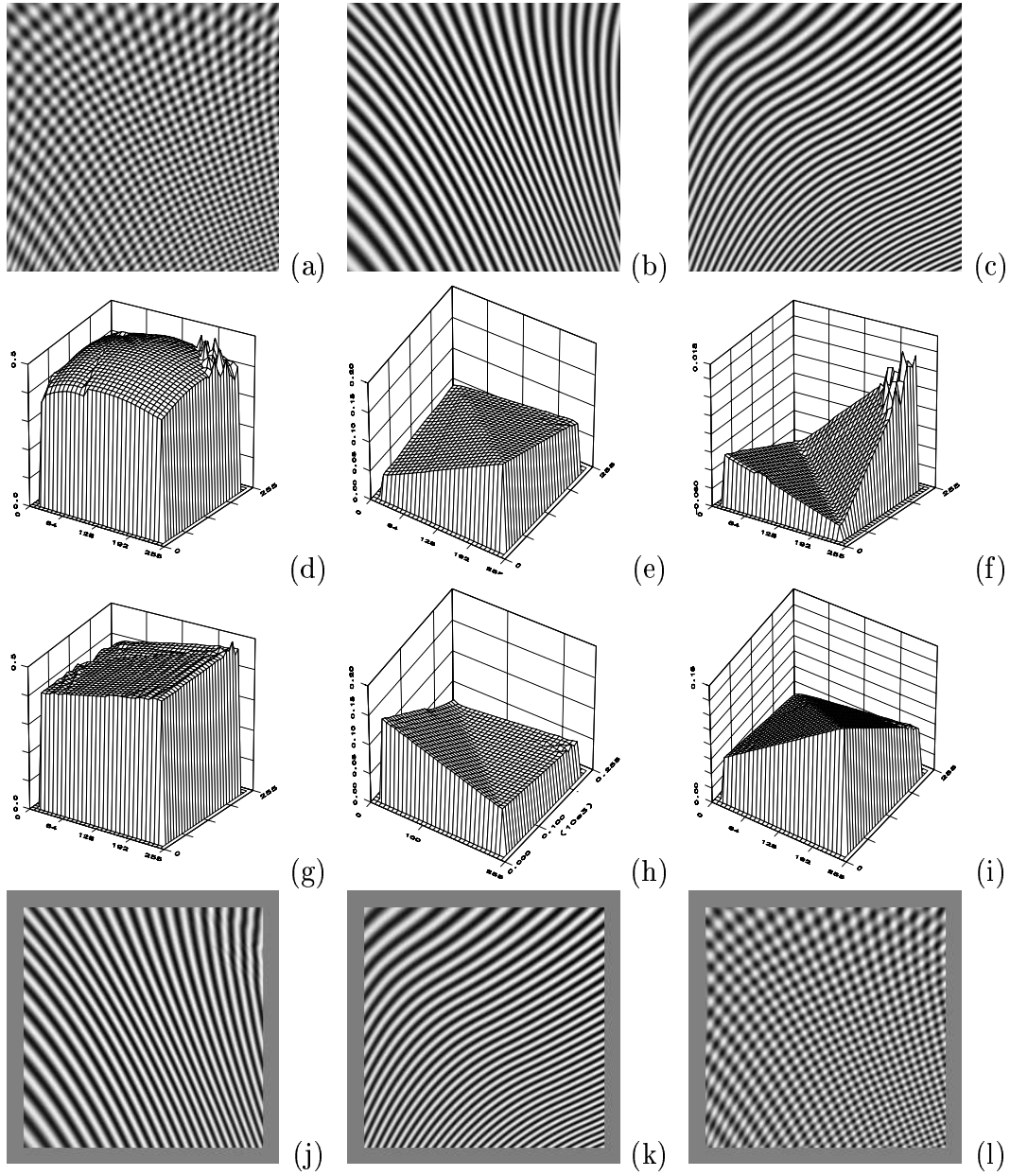


Figure 8: Multi-component AM-FM representation and reconstruction of a synthetic image. (a) Nonstationary two-component image. (b) True values for component one. (c) True values for component two. (d) Amplitude estimates for component one. (e) Horizontal frequency estimates for component one. (f) Vertical frequency estimates for component one. (g) Amplitude estimates for component two. (h) Horizontal frequency estimates for component two. (i) Vertical frequency estimates for component two. (j) Reconstruction of component one. (k) Reconstruction of component two. (l) Reconstructed image.

information content of the image. Most pixels in the reconstructed image are covered by *three or fewer* reconstructed AM-FM components.

6 Conclusion & Future Work

In this report we introduced the multi-component AM-FM representation as a powerful new technique for modeling, analysis, representation, and analysis of images. The representation efficiently characterizes nonstationary, multi-partite images as sums of locally-coherent, nonlinear AM-FM functions that naturally and elegantly facilitate analysis in terms of the nonstationarities, where important visual and perceptual cues are often manifest. The 2D discrete Fourier transform is, in fact, the simplest multi-component AM-FM representation. For any image, more general multi-component AM-FM representations always exist, and these may be both more efficient than the DFT and better suited for analysis than the DFT.

Using a statistical state-space image component model, we developed practical techniques for computing the multi-component AM-FM representation and for recovering the image from the computed representation. Under the assumptions that the multiband 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 multi-component AM-FM representations of two multi-partite images, and obtained reconstructions in remarkable agreement with the original images.

Important future work remaining in this area includes overcoming the extremely difficult problems in treating complicated natural images, which may contain many components closely spaced in frequency and supported on irregularly shaped regions of the image.

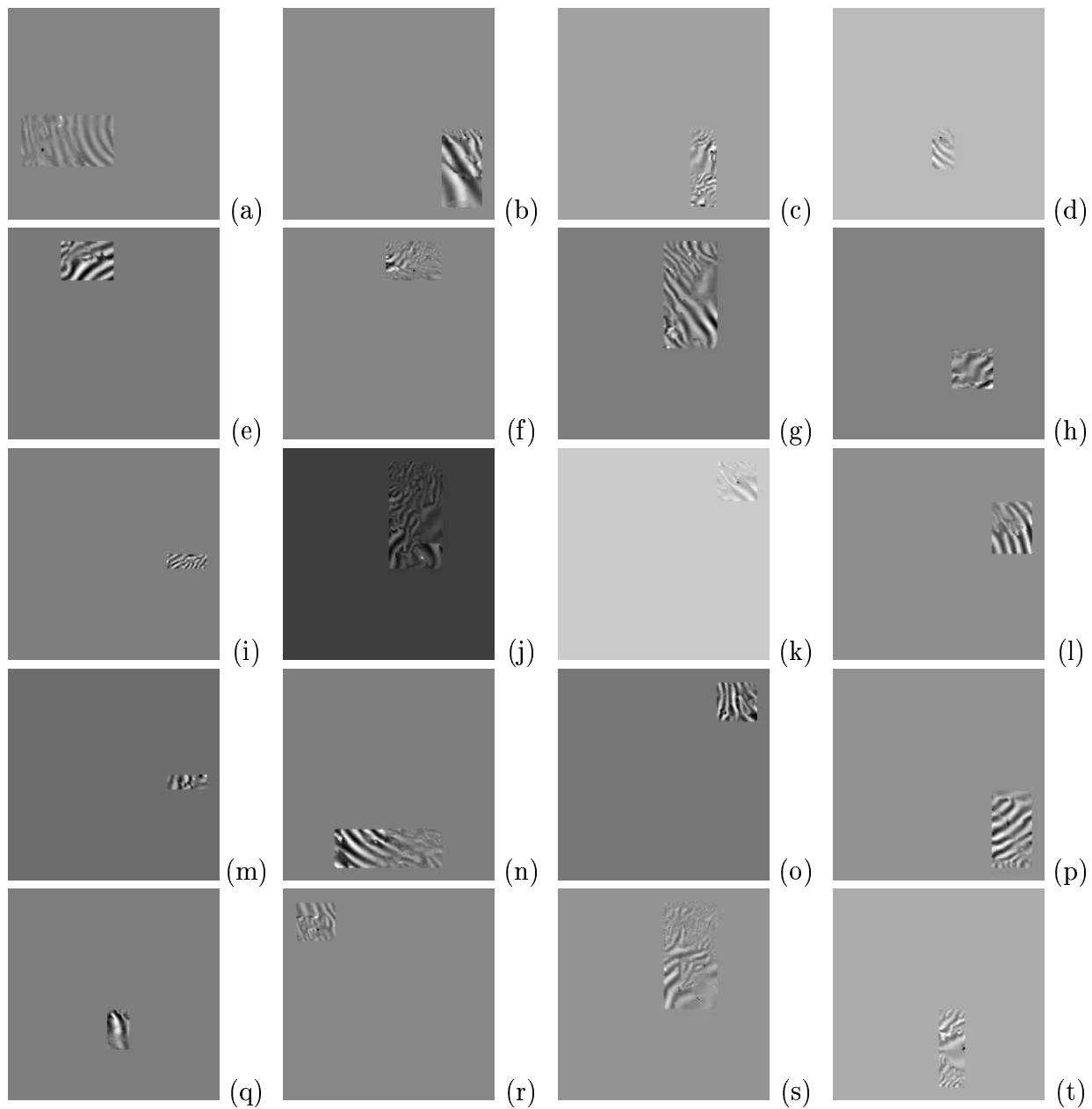


Figure 9: Reconstructions of 20 AM-FM components of the tree image from their amplitude and frequency estimates.

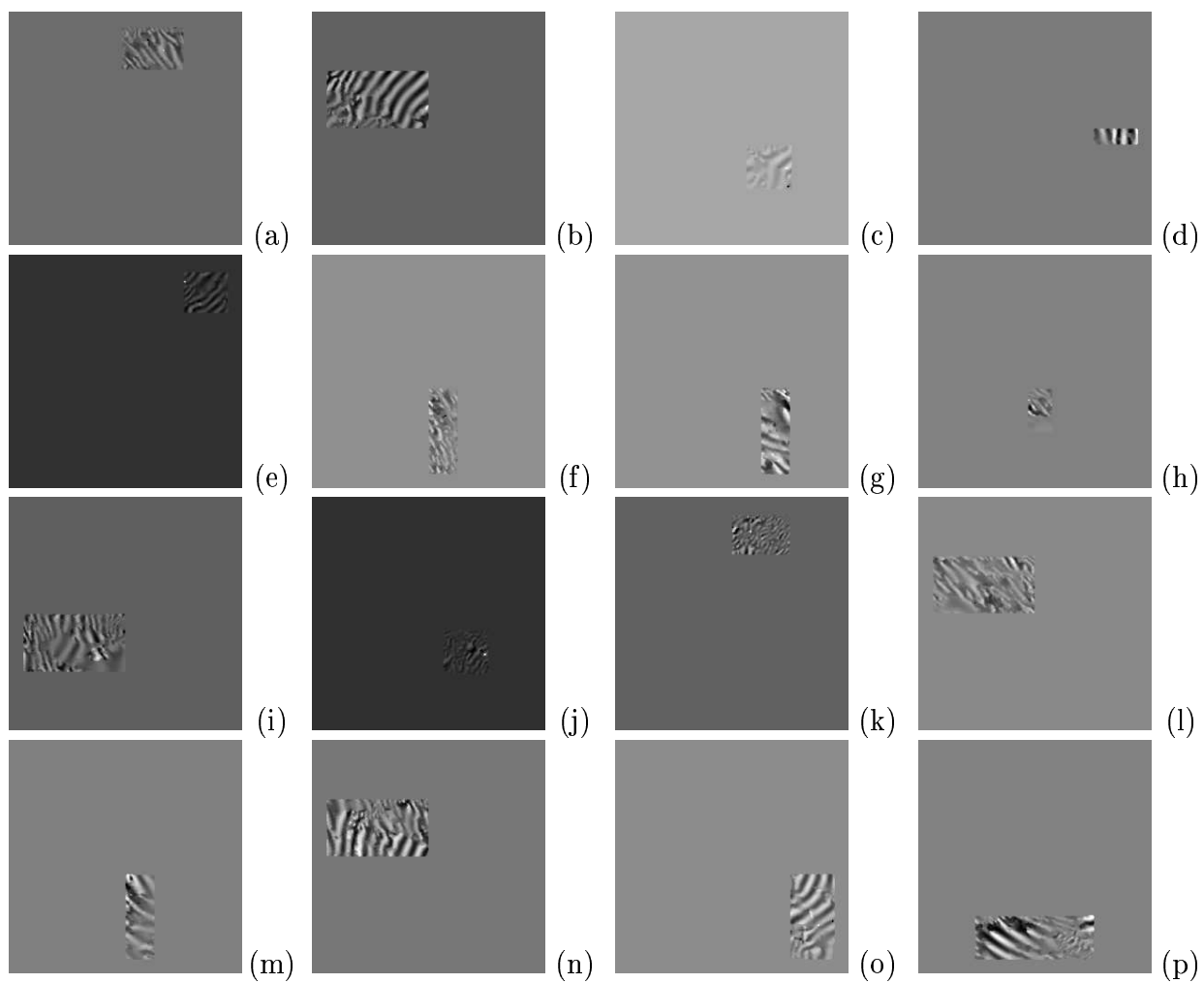


Figure 10: Reconstructions of 16 AM-FM components of the tree image from their amplitude and frequency estimates.

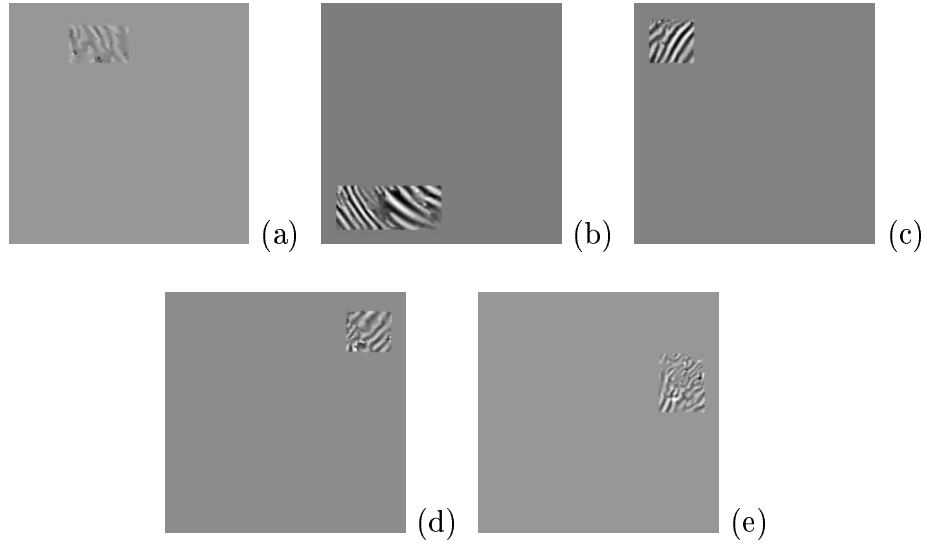


Figure 11: Reconstructions of 5 AM-FM components of the tree image from their amplitude and frequency estimates.

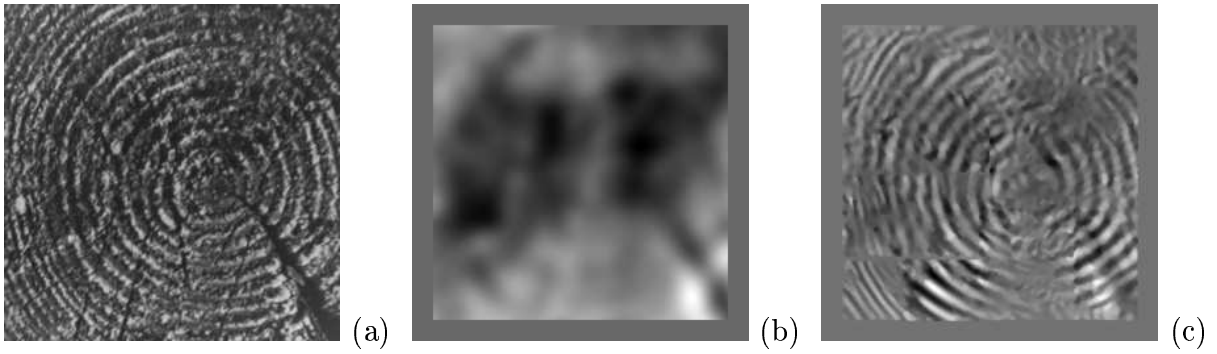


Figure 12: Reconstruction of the complicated natural multi-partite Brodatz texture image *tree* from 41 computed AM-FM components and a low-frequency component. (a) Tree image. (b) Low-pass component extracted by linear filtering. (c) Reconstruction from 41 AM-FM components and the low-pass component.

References

- [1] J. P. Havlicek and A. C. Bovik, “AM-FM models, the analytic image, and nonlinear demodulation techniques”, Tech. Rept. TR-95-001, Center for Vision and Image Sciences, The University of Texas at Austin, March 1995.
- [2] R. Bajcsy and L. Lieberman, “Texture gradients as a depth cue”, *Comput. Graphics Image Process.*, vol. 5, pp. 52–67, 1976.
- [3] M. Kass and A. Witkin, “Analyzing oriented patterns”, *Comput. Graphics Image Process.*, vol. 37, pp. 362–385, 1987.
- [4] A. C. Bovik, M. Clark, and W. S. Geisler, “Multichannel texture analysis using localized spatial filters”, *IEEE. Trans. Pattern Anal. Machine Intell.*, vol. PAMI-12, no. 1, pp. 55–73, January 1990.
- [5] A. C. Bovik, “Analysis of multichannel narrow-band filters for image texture segmentation”, *IEEE. Trans. Signal Proc.*, vol. SP-39, no. 9, pp. 2025–2043, September 1991.
- [6] A. C. Bovik, “Variational pattern analysis using Gabor wavelets”, in *Proc. IEEE Int’l. Conf. Acoust., Speech, Signal Proc.*, San Francisco, CA, March 1992, pp. IV669 – IV672.
- [7] B. J. Super and A. C. Bovik, “Solution to shape-from-texture by wavelet-based measurement of local spectral moments”, Tech. Rept. TR-92-5-80, Computer and Vision Research Center, The University of Texas at Austin, November 1991.
- [8] T. Y. Chen, A. C. Bovik, and B. J. Super, “Multiscale stereopsis via gabor filter phase response”, in *Proc. IEEE Int’l. Conf. Syst., Man, and Cyber.*, SanAntonio, TX, October 2 - 5 1994, pp. 55 – 60.
- [9] J. P. Havlicek, D. S. Harding, and A. C. Bovik, “Multi-component signal demodulation and reconstruction using AM-FM modulation models”, in *Proc. 1995 IEEE Workshop Nonlin. Signal and Image Proc.*, Neos Marmaras, Halkidiki, Greece, June 20 - 22 1995.
- [10] P. Maragos and A. C. Bovik, “Demodulation of images modeled by amplitude-frequency modulation using multidimensional energy separation”, in *Proc. IEEE Int’l. Conf. Image Proc.*, Austin, TX, 1994, pp. III421–III425.
- [11] H. Knutsson, C.F. Westin, and G. Granlund, “Local multiscale frequency and bandwidth estimation”, in *Proc. IEEE Int’l. Conf. Image Proc.*, Austin, TX, 1994, pp. I36–I40.
- [12] J. P. Havlicek and A. C. Bovik, “Multi-component AM-FM image models and wavelet-based demodulation with component tracking”, in *Proc. IEEE Int’l. Conf. Image Proc.*, Austin, TX, November 13-16 1994, pp. I41 – I45.
- [13] P. Maragos, J. F. Kaiser, and T. F. Quatieri, “On amplitude and frequency demodulation using energy operators”, *IEEE Trans. Signal Proc.*, vol. SP-41, no. 4, pp. 1532–1550, April 1993.

- [14] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with applications to speech analysis", *IEEE. Trans. Signal Proc.*, vol. SP-41, no. 10, pp. 3024–3051, October 1993.
- [15] A. C. Bovik, P. Maragos, and T. F. Quatieri, "AM-FM energy detection and separation in noise using multiband energy operators", *IEEE. Trans. Signal Proc.*, vol. SP-41, no. 12, pp. 3245–3265, December 1993.
- [16] A. C. Bovik, P. Maragos, and T. F. Quatieri, "Demodulation of AM-FM signals in noise using multiband energy operators", in *Proc. IEEE Int'l. Symp. Info. Theory*, San Antonio, TX, January 17-22 1993.
- [17] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "On separating amplitude from frequency modulations using energy operators", in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Proc.*, San Francisco, CA, March 1992, pp. II1–II4.
- [18] 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, October 26 - 28 1992, pp. 805–810.
- [19] A. C. Bovik, P. Maragos, and T. F. Quatieri, "Measuring amplitude and frequency modulations in noise using multiband energy operators", in *Proc. IEEE Int'l. Symp. Time-Frequency and Time-Scale Analysis*, Victoria, B.C., Canada, October 4-6 1992.
- [20] P. Maragos, A. C. Bovik, and T. F. Quatieri, "A multidimensional energy operator for image processing", in *Proc. SPIE Symp. Visual Commun. Image Proc.*, Boston, MA, November 16-18 1992, pp. 177–186.
- [21] P. Maragos and A. C. Bovik, "Image demodulation using multidimensional energy separation", *J. Opt. Soc. Amer. A*, vol. 12, no. 9, pp. 1867–1876, September 1995.
- [22] P. Maragos, T. F. Quatieri, and J. F. Kaiser, "Speech nonlinearities, modulations, and energy operators", in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Proc.*, Toronto, Ont., Canada, May 1991, pp. 421–424.
- [23] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis", Tech. Rept. 91-17, Harvard Robotics Laboratory, Harvard University, November 1991.
- [24] L. Cohen, *Time-Frequency Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1995.
- [25] L. Cohen, "Time-frequency distributions – a review", *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, July 1989.
- [26] L. Cohen and C. Lee, "Standard deviation of instantaneous frequency", in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Proc.*, Glasgow, Scotland, May 23-26 1989.
- [27] L. Cohen and C. Lee, "Instantaneous bandwidth for signals and spectrogram", in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Proc.*, Albuquerque, NM, April 1990, pp. 2451–2454.

- [28] A. Papoulis, *Systems and Transforms with Applications in Optics*, McGraw-Hill, New York, 1968.
- [29] B. Boashash, G. Jones, and P. O'Shea, "Instantaneous frequency of signals: Concepts, estimation techniques, and applications", *SPIE Advanced Algorithms and Architectures for Signal Proc.*, vol. 1152, pp. 382–400, 1989.
- [30] A. Reilly and B. Boashash, "Comparison of time-frequency signal analysis techniques with application to speech recognition", *SPIE Advanced Algorithms and Architectures for Signal Proc.*, vol. 1770, pp. 339–350, 1992.
- [31] A. Reilly, G. Frazer, and B. Boashash, "Analytic signal generation – tips and traps", *IEEE. Trans. Signal Proc.*, vol. SP-42, no. 11, pp. 3241–3245, November 1994.
- [32] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, NJ, 1970.
- [33] Alexandros Potamianos and Petros Maragos, "A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation", Tech. Rept. 92-8, Harvard Robotics Laboratory, Harvard University, July 1992.
- [34] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo, "Localized measurement of emergent image frequencies by Gabor wavelets", *IEEE. Trans. Info. Theory*, vol. IT-38, no. 2, pp. 691–712, March 1992.
- [35] J. P. Havlicek, A. C. Bovik, M. D. Desai, and D. S. Harding, "The discrete quasi-eigenfunction approximation", in *Proc. Int'l. Conf. on Digital Signal Proc.*, Limassol, Cyprus, June 26 - 28 1995.
- [36] A. C. Bovik, J. P. Havlicek, and M. D. Desai, "Theorems for discrete filtered modulated signals", in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Proc.*, Minneapolis, MN, April 27 - 30 1993, pp. III153–III156.
- [37] A. C. Bovik, "A bound involving n-dimensional instantaneous frequency", *IEEE. Trans. Circ. Syst.*, vol. CAS-38, no. 11, pp. 1389–1390, November 1991.
- [38] J. E. Marsden and A. J. Tromba, *Vector Calculus*, W. H. Freeman, San Francisco, CA, 1981.
- [39] E. Vanmarcke, *Random Fields, Analysis and Synthesis*, MIT Press, Cambridge, MA, 1983.
- [40] A. C. Bovik, J. P. Havlicek, D. S. Harding, and M. D. Desai, "Limits on discrete modulated signals", *IEEE Trans. Signal Proc.*, *submitted*.