MULTI-COMPONENT AM-FM IMAGE MODELS AND WAVELET-BASED DEMODULATION WITH COMPONENT TRACKING

Joseph P. Havlicek and Alan C. Bovik

Laboratory for Vision Systems, University of Texas, Austin, TX 78712-1084

ABSTRACT

In this paper we utilize multi-component AM-FM functions to model multi-partite nonstationary images that are locally coherent, yet globally wideband. We also detail an approach for simultaneously estimating the modulating functions associated with each of the multiple components. Components are isolated by a multiband bank of Gabor wavelets, and estimates of the modulating functions are derived from each channel. We use simple one-dimensional Kalman filters to track each identified component across the channels.

1. INTRODUCTION

We explore new techniques for extracting multiple amplitude modulation (AM) and frequency modulation (FM) components from images. For the first time, we demonstrate effective identification, demodulation, and tracking of multiple globally wideband, nonstationary yet locally coherent individual signal components making up a multi-component image.

AM-FM modeling techniques and associated demodulation algorithms have been the subject of significant recent interest [3] - [11]. Bovik, *et. al.*, demonstrated powerful characterizations of texture in terms of *emergent frequency*, or frequency modulation models [3]. Maragos, Kaiser, and Quatieri [6] - [10], and Bovik, Maragos, and Quatieri [9], [11] characterized and popularized the *Teager-Kaiser Energy Operator* for AM-FM signal demodulation. A multi-dimensional version of the Teager-Kaiser operator and its associated *Energy Separation Algorithm* was developed by Maragos, Bovik, and Quatieri [4].

Havlicek, Bovik, and Maragos introduced a related demodulation algorithm unique in its ability to estimate multidimensional instantaneous frequency with correct sign information, and used it to demodulate the dominant components of multi-component images [5]. In multiple dimensions, the *signs* of the instantaneous frequency components play an important role in determining the *local orientation* of the instantaneous frequency vector. These signs definitely are significant in representing orientations *outside* of the first quadrant of the frequency hyperplane.

In the present paper, we combine this demodulation algorithm with computationally efficient low-order 1-D Kalman filters to effectively track and demodulate the individual components of an n-dimensional multi-component image modeled as

$$t(\mathbf{x}) = \sum_{k=1}^{K} a_k(\mathbf{x}) \exp[j\varphi_k(\mathbf{x})], \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $t : \mathbb{R}^n \to \mathbb{C}$, $a : \mathbb{R}^n \to [0, 1]$, and $\varphi : \mathbb{R}^n \to \mathbb{R}$.

2. SINGLE COMPONENT CASE

For a single nonstationary, yet locally coherent n-dimensional complex valued image component modeled as

$$t(\mathbf{x}) = a(\mathbf{x}) \exp[j\varphi(\mathbf{x})], \tag{2}$$

which is "locally narrowband" in the sense that the amplitude $a(\mathbf{x})$ and instantaneous frequency $\nabla\varphi(\mathbf{x})$ do not vary too rapidly, the demodulation problem involves estimating the amplitude envelope $a(\mathbf{x})$ and the instantaneous frequencies $\nabla\varphi(\mathbf{x})$ that characterize the local image structure. We accomplish this by observing that

$$\nabla t(\mathbf{x}) = jt(\mathbf{x})\nabla\varphi(\mathbf{x}) + \exp[j\varphi(\mathbf{x})]\nabla a(\mathbf{x}). \tag{3}$$

Hence

$$\frac{\nabla t(\mathbf{x})}{jt(\mathbf{x})} = \nabla\varphi(\mathbf{x}) - j\frac{\nabla a(\mathbf{x})}{a(\mathbf{x})}. \tag{4}$$

This leads immediately to the demodulation algorithm

$$|t(\mathbf{x})| = |a(\mathbf{x}) \exp[j\varphi(\mathbf{x})]| = |a(\mathbf{x})| = a(\mathbf{x}) \tag{5}$$

and

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

The quantity

$$B(\mathbf{x}) = \frac{|\nabla a(\mathbf{x})|^2}{|a(\mathbf{x})|^2} \tag{7}$$

is the n-dimensional extension of the *instantaneous bandwidth*, defined in one dimension by Cohen [12]. This quantity is related to the spread of frequencies present around the dominant, or *emergent*, instantaneous frequency of a component on a localized basis.

In (5), (6), if the image of interest is real, we form its complex extension by adding j times its Hilbert transform prior to demodulation. As we shall point out in Section 3, multiband wavelet processing is required to isolate individual components in the case of a multiple component signal,

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.

and hence there is no computational penalty in using the Hilbert transform extension method. In fact, it saves us from having to compute half of the multiplies in all subsequent FFT-based filtering operations.

3. MULTI-COMPONENT CASE

Since the demodulation algorithm (6) is nonlinear, it breaks down in the presence of out-of-band additive noise, when multiple locally narrowband signal components are present, and when both multiple components and broadband noise are present due to interference between the various signal components [11]. Therefore, in these situations it becomes necessary to *isolate* the various locally narrowband components prior to demodulation. We accomplish this by passing the multi-component image through a multiple bandpass (multiband) bank of Gabor wavelets, which are optimal in their conjoint time-frequency uncertainty [2]. This scheme isolates the AM-FM components on a spatially local basis, and we subsequently apply the demodulation algorithm (5), (6) to the filtered channel outputs with the slight modification that the amplitude estimates so obtained must be divided by the appropriate channel filter magnitude evaluated at the estimated instantaneous frequencies. Although this *filtered* demodulation scheme is an approximation, we have theoretical results which tightly bound the approximation error, and by which we can force the error to be small by designing the multiband filters prudently [5].

3.1. Gabor Wavelet Filter Bank in 2D

All of the filters have unit L^2 norm in both domains. The equation for a filter with center frequency (u_m, v_m) is

$$g_m(x_1,x_2) = \frac{1}{\sigma_m\sqrt{2\pi}} \exp\left[\frac{-1}{4\sigma_m^2}(x_1^2+x_2^2)\right] \exp\left[j2\pi(u_m x_1 + v_m x_2)\right],$$

with Fourier transform

$$G_m(u,v) = 2\sigma_m\sqrt{2\pi}\exp\left\{-4\pi^2\sigma_m^2 \left[(u-u_m)^2+(v-v_m)^2\right]\right\}.$$

Along rays, the filter radial center frequencies progress geometrically with common ratio R, such that the radial center frequencies of adjacent filters are related by $r_{m+1} = Rr_m$. With $\gamma = (2^B-1)^2/(2^B+1)^2$, we take $\sigma_m = \sqrt{-\ln\eta/(4\pi^2 r_m^2\gamma)}$, so that each filter has a radial η-peak bandwidth of B octaves and an orientation bandwidth $\Theta = 2\arctan\sqrt{\gamma}$. Moreover, by spacing adjacent rays at angles of

$$\Lambda = 2\arcsin\sqrt{\frac{1}{4R}[(R^2+1)(\gamma-1)+2R(\gamma+1)]},$$

we ensure that any four adjacent filters intersect precisely where each is at a fraction η of peak response. With this dense spacing, every point in the frequency plane is covered by a filter responding at η-peak or higher.

For the examples in this paper, we used $R = 1.8$, $B = 1$, and $\eta = \frac{1}{2}$, giving eight rays in the right-half plane. The radial center frequency of the first filter on each ray was 9.6 cycles per image. Each response image was postfiltered as described in [1].

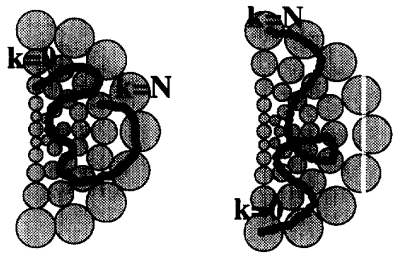

Figure 1: Frequency tracks of two image components in a region. The index k advances as we traverse the image in the spatial domain.

3.2. Multi-Component Algorithm

With the filtered demodulation algorithm, every channel in the multiband filter bank produces observations of $a(\mathbf{x})$ and of $\nabla\varphi(\mathbf{x})$ at every pixel in the image. The problem of multi-component demodulation then becomes one of determining which channel should be used in estimating the modulating functions of *each* component at each pixel. Figure 1 depicts a case where two components exist in an image region containing $N+1$ pixels. Suppose that we specify an order for the pixels such that traversing them in order results in our following a contiguous path in the spatial domain. Then, as we traverse the pixels in order, the pixel number k varies from 0 to N. With this variation in k, the instantaneous frequency of each component also maps out a track in the frequency domain, as depicted in Figure 1. At each pixel, we decide which channel to use in estimating the modulating functions of a given component by following that component's frequency domain track with a track processor. A block diagram of the overall scheme is shown in Figure 2. The demodulation algorithm (5), (6) is performed in the blocks marked "ESA".

3.3. Component Tracking

In this section we discuss the details of a track processor which is both simple and computationally efficient. For each image component, the quantities which we track are, with reference to eq. 1, a_k and each component of $\nabla\varphi_k$. In two dimensions, we let $U(\mathbf{x}) = \frac{\partial}{\partial x_1}\varphi(\mathbf{x})$ and $V(\mathbf{x}) = \frac{\partial}{\partial x_2}\varphi(\mathbf{x})$.

We make no specific assumptions about statistical correlation between these quantities, and hence track each one separately. For the remainder of this development, we refer to an arbitrary one of them as $f(x_1,x_2)$. We map the discrete domain vector function $f(m,n)$ to a one dimensional function f_k by

$$k = \begin{cases} mN+n, & \text{m even} \\ (m+1)N-n-1, & \text{m odd,} \end{cases} \quad (8)$$

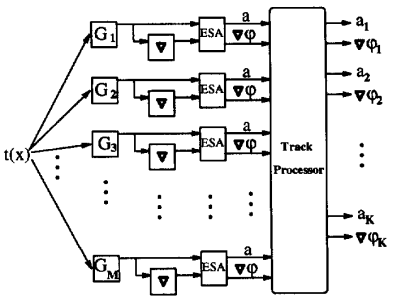

Figure 2: Block diagram of the multi-component algorithm.

where N is the row dimension of the image (assumed square), and write the function f_k in a first order Taylor series:

$$f_k = f_{k-1} + f'_{k-1} + \int_{k-1}^{k} (k-t) f''(t) dt, \qquad (9)$$

which is exact up to sampling errors. Since we do not expect f_k to vary too rapidly or too wildly from pixel to pixel, we do not model the higher order terms of the series explicitly, but rather consider that f' is essentially a constant plus a drift, where the drift is an uncorrelated noise process u_k. This immediately gives us the simple and widely applicable plant model

$$\begin{bmatrix} f_{k+1} \\ f'_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} f_k \\ f'_k \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_k. \qquad (10)$$

Although this model is perfectly valid, it should not be expected to lead to the best possible track filter since it does not make use of all of f_k's two-dimensional neighbors. Nevertheless, the design of an elegant Kalman track filter involving only *scalar* equations is straightforward from the model (10):

$$\widehat{f}_{k|k} = \widehat{f}_{k|k-1} + \alpha_k (f_k^o - \widehat{f}_{k|k-1}) \qquad (11)$$

$$\widehat{f'}_{k+1|k} = \widehat{f'}_{k|k-1} + \beta_k (f_k^o - \widehat{f}_{k|k-1}) \qquad (12)$$

$$\widehat{f}_{k+1|k} = \widehat{f}_{k|k} + \widehat{f'}_{k+1|k}, \qquad (13)$$

where α_k and β_k are the Kalman gain sequences, f_k^o is the observation of f at pixel k, $\widehat{f}_{k|k-1}$ is the prediction of f at pixel k given $k-1$ observations, and $\widehat{f}_{k|k}$ is the estimate of f at pixel k given k observations.

3.3.1. Track Gating

At each pixel, let a_n^o be the amplitude estimate obtained by applying (5) to channel n, and let U_n^o, V_n^o be the instantaneous frequency estimates obtained by applying (6) to channel n. Let $\widehat{a}_m$ be the predicted amplitude of track m at the pixel, obtained from (13), and let $\widehat{U}_m$ and $\widehat{V}_m$ be the predicted instantaneous frequencies.

For each track, let $\widetilde{e}_a$ be the smoothed residual in amplitude, and $\widetilde{e}_U$, $\widetilde{e}_V$ be the smoothed residuals in instantaneous frequency, where the smoothing is performed with a single-pole IIR filter. We define the radius of the track gate for track m at a pixel by

$$\mathcal{R}_m = \sqrt{\tau(5\widetilde{e}_a)^2 + \lambda(5\widetilde{e}_U)^2 + \zeta(5\widetilde{e}_V)^2},$$

where $\tau \propto |\widehat{a}_m|^{-2}$, $\lambda \propto |\widehat{U}_m|^{-2}$, and $\zeta \propto |\widehat{V}_m|^{-2}$, and the distance from track m to the estimates of channel n by

$$\varepsilon_{m,n} = \sqrt{\tau(\widehat{a}_m - a_n^o)^2 + \lambda(\widehat{U}_m - U_n^o)^2 + \zeta(\widehat{V}_m - V_n^o)^2}.$$

We say that the estimates of channel n lie ***within*** the gate of track m at a particular pixel if $\varepsilon_{m,n} \leq \mathcal{R}_m$.

3.3.2. New Track Starts

We would like to try starting a new track on a channel that produces a large amplitude observation a_n^o, and whose frequency observations U_n^o, V_n^o do not lie within the gate of an existing track. However, our confidence in the observations produced by a channel is highest when the observed frequencies are close to the channel center frequency, since this generally affords us improved immunity against out-of-band information through an enhanced SNR. The quantity $|G_n(u,v)| \,/\, \max_{u,v} |G_n(u,v)|$ lies between zero and one, and increases as (u,v) moves closer to the filter center frequency. Hence we define the *quality* of channel n at each pixel by

$$Q_n = \frac{a_n^o \, |G_n(U_n^o, V_n^o)|}{\max_{u,v} |G_n(u,v)|}. \qquad (14)$$

At each pixel, if the observations from the channels with the highest values of Q_n do not lie within the gates of any existing tracks, then we start new tracks on those channels.

3.3.3. Track Updating

At each pixel, we update each existing track using the observations from the channel whose center frequency is closest to the predicted frequencies of the track, provided that these observations lie within the track gate. If the observations are outside the gate, however, we coast the track by updating it directly from the predictions.

3.3.4. Vertical Updating

We mentioned in Section 3.3 that a disadvantage of the one-dimensional plant model is the fact that tracks are updated at a pixel using only information from horizontally neighboring pixels. We partially ameliorate this situation by saving the state of each track at every pixel. When track processing reaches the corresponding pixel on the next image row, if the observations lie within the gate of the saved *vertical* state, and the if distance $\varepsilon_{m,n}$ from the saved state is smaller than the distance to the current (horizontal) state of the track, then we restore the track to the saved state. This has the effect of creating a track split at the pixel on the previous row, with one of the split tracks continuing horizontally while the other propagates vertically for one row.

4. EXAMPLES

The results of applying the multi-component demodulation algorithm with component tracking described in Sections 2 - 3 to track two components of a multi-component tree image are shown in Figure 3. All images are scaled for maximum dynamic range.

Figure 3(a) shows the image. It's discrete Hilbert transform is shown in Figure 3(b), and we formed the associated analytic image by adding j times the Hilbert transform to the image. The amplitudes of the two tracked components are shown in Figures 3(c) and (f), respectively. The demodulated amplitude tends to be large in regions where the image component exhibits high *contrast*.

The instantaneous bandwidths, computed from the two demodulated amplitudes, are shown in Figures 3(d) and (e). The instantaneous bandwidth is smallest at points where only a single instantaneous frequency is emergent, while it tends to be large in regions characterized by the presence of numerous subemergent frequencies. In an imprecise sense, our confidence in the estimated amplitudes and frequencies is best where the instantaneous bandwidth is small, and worst where it is large. This generalization can be made rigorous using the main theorem of [5].

Finally, the estimates of the instantaneous frequency vectors $\nabla\varphi_k(\mathbf{x})$ for each tracked component are shown in the needle diagrams of Figures 3(e) and (f). Only one needle is shown for each block of 8×8 pixels. The needle orientation is the pointwise estimate of the orientation of $\nabla\varphi_k(\mathbf{x})$, which is generally normal to visually perceived contours in the image. The length of each needle is *inversely* proportional to $|\nabla\varphi_k(\mathbf{x})|$, which makes it proportional to the *instantaneous period*. This display convention facilitates visual comparison of the frequency estimates with the original image, since spatially extended image features characterized by the presence of low frequencies are associated with long needles in the needle diagram, while smaller features characterized by the presence of high frequencies are associated with shorter needles.

Clearly, the image in Figure 3(a) contains additional significant components in addition to the two we have tracked; determination of *how many* components are actually present remains an open problem.

In Figure 4, we apply the approach to another interesting image which displays nonstationarity with local coherence. Again, all images are scaled for maximum dynamic range. The original 256×256 burlap image is shown in Figure 4(a), and the Hilbert transform is shown in Figure 4(b). We tracked three components, and the estimated amplitudes are shown in Figures 4(c), (f), and (i). Notice that tracking of the third component did not start until row 16 of the image, hence the amplitude estimate is zero for rows zero to 15 (and the instantaneous bandwidth is 255 in this region). The instantaneous bandwidth images are shown in Figures 4(d), (g), and (j), while needle diagrams of the estimated instantaneous frequency vectors are shown in Figures 4(e), (h), and (k). Once again, the image clearly contains additional important information that we have not captured in tracking only three components.

5. FUTURE WORK

Important future work in this area includes development of true multi-dimensional track filters and reconstruction of an image from the estimated modulating signals, including determining how many components are present.

6. REFERENCES

[1] A.C. Bovik, "Analysis of multichannel narrow-band filters for image texture segmentation," *IEEE Trans. Signal Process.*, vol. SP-39, pp. 2025-2043, Sept. 1991.

[2] 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, pp. 55-73, Jan. 1990.

[3] A.C. Bovik, N. Gopal, T. Emmoth and A. Restrepo, "Localized measurement of emergent image frequencies by Gabor wavelets," *IEEE Trans. Info. Theory*, Special Issue on Wavelet Transforms and Multiresolution Signal Analysis, vol. IT-38, pp. 691-712, Mar. 1992.

[4] P. Maragos, A.C. Bovik and T.F. Quatieri , "A multidimensional energy operator for image processing," in *Proc. SPIE Symp. Visual Commun. Image Process.*, Boston, MA, Nov. 1992.

[5] J.P. Havlicek, A.C. Bovik, and P. Maragos, "Modulation models for image processing and wavelet-based image demodulation," *Ann. Asilomar Conf. Signals, Syst., Comput.*, Montery, CA, Oct. 1992.

[6] P. Maragos, T.F. Quatieri and J.F. Kaiser, "Speech nonlinearities, modulations, and energy operators," in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Process.*, Toronto, Ontario, Canada, May 1991.

[7] P. Maragos, T.F. Quatieri and J.F. Kaiser, "On separating amplitude from frequency modulations using energy operators," in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Process.*, San Francisco, CA, May 1992.

[8] P. Maragos, J.F. Kaiser and T.F. Quatieri , "On amplitude and frequency demodulation using energy operators," *IEEE Trans. Signal Process.*, vol. SP-41, pp. 1532 - 1550, Apr. 1993.

[9] A.C. Bovik, P. Maragos and T.F. Quatieri, "Measuring amplitude and frequency modulations in noise using multiband energy operators," *IEEE Int'l. Symp. Time-Freq. and Time-Scale Anal.*, Victoria, British Columbia, Canada, Oct. 1992.

[10] P. Maragos, J.F. Kaiser, and T.F. Quatieri, "Energy separation in signal modulations with applications to speech analysis," *IEEE Trans. Signal Process.*, vol. SP-41, pp. 3024 - 3051, Oct. 1993.

[11] A.C. Bovik, P. Maragos and T.F. Quatieri, "AM-FM energy detection and separation in noise using multiband energy operators," *IEEE Trans. Signal Process.*, Special Issue on Wavelets and Signal Processing, vol. SP-41, pp. 3245 - 3265, Dec. 1993.

[12] L. Cohen and C. Lee, "Instantaneous bandwidth for signals and spectrogram," in *Proc. IEEE Int'l. Conf. Acoust., Speech, Signal Process.*, Albuquerque, NM, Apr. 1990.

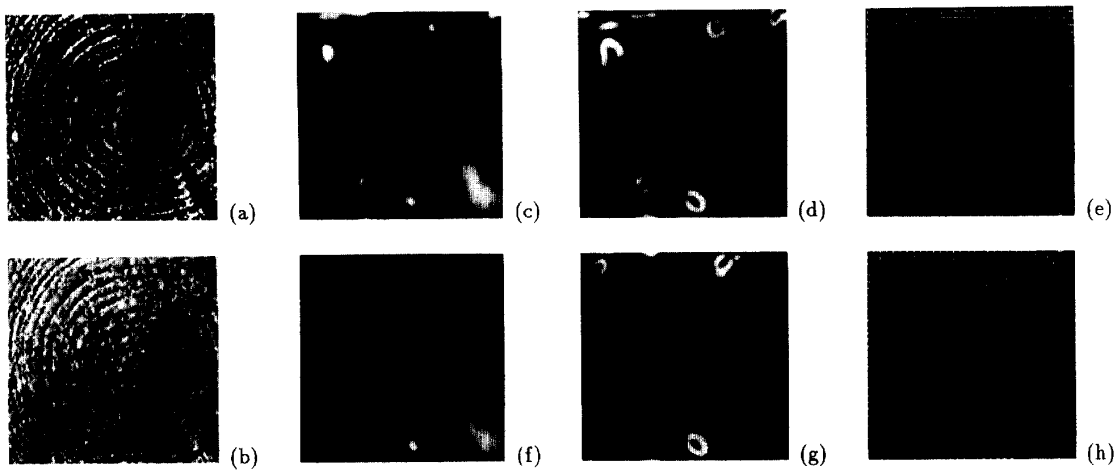

Figure 3: (a) Tree image, real part. (b) Imaginary part, computed by Hilbert transform. (c) Estimate of $a_1(\mathbf{x})$. (d) Estimated instantaneous bandwidth of component one, $|\nabla a_1(\mathbf{x})|/|a_1(\mathbf{x})|$. (e) Needle diagram showing estimate of $\nabla\varphi_1(\mathbf{x})$. The length of each needle is inversely proportional to to $|\nabla\varphi_1(\mathbf{x})|$ (proportional to wavelength). (f) Estimate of $a_2(\mathbf{x})$. (g) Estimate of $|\nabla a_2(\mathbf{x})|/|a_2(\mathbf{x})|$. (h) Estimate of $\nabla\varphi_2(\mathbf{x})$.

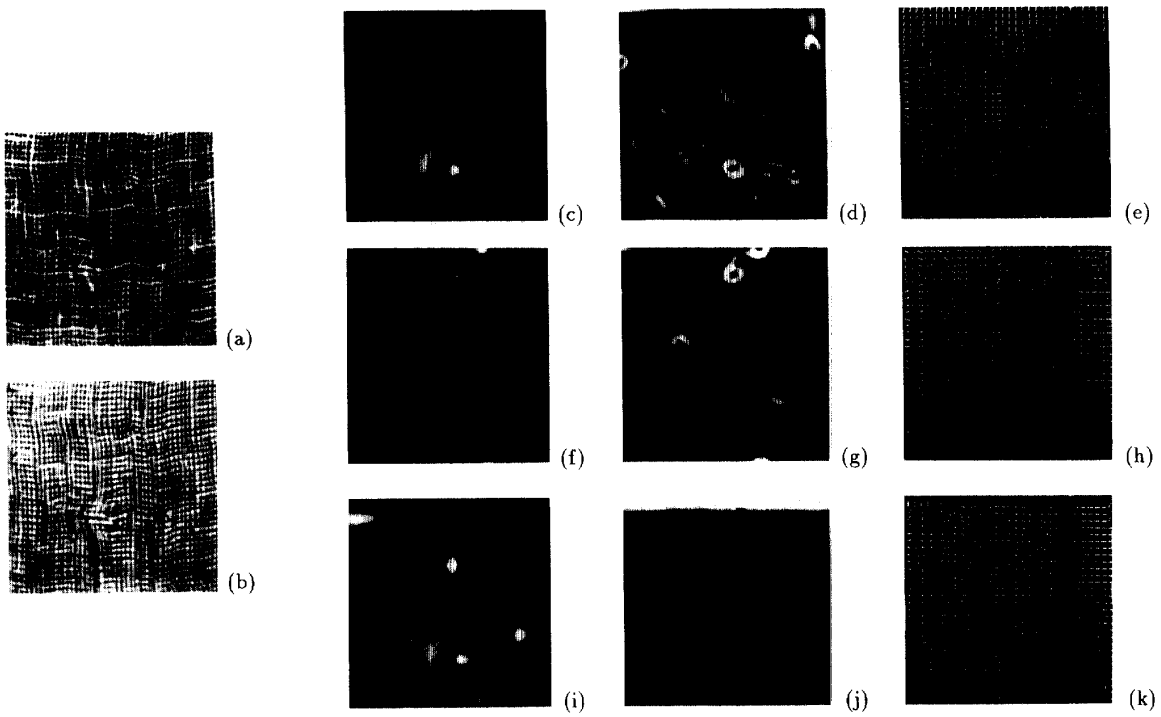

Figure 4: (a) Burlap image, real part. (b) Imaginary part, by Hilbert transform. (c) Estimate of $a_1(\mathbf{x})$. (d) Estimate of $|\nabla a_1(\mathbf{x})|/|a_1(\mathbf{x})|$. (e) Estimate of $\nabla\varphi_1(\mathbf{x})$. (f) Estimate of $a_2(\mathbf{x})$. (g) Estimate of $|\nabla a_2(\mathbf{x})|/|a_2(\mathbf{x})|$. (h) Estimate of $\nabla\varphi_2(\mathbf{x})$. (f) Estimate of $a_3(\mathbf{x})$. (g) Estimate of $|\nabla a_3(\mathbf{x})|/|a_3(\mathbf{x})|$. (h) Estimate of $\nabla\varphi_3(\mathbf{x})$. Tracking of component 3 did not begin until row 16.