Extracting Essential Modulated Image Structure

Joseph P. Havlicek, David S. Harding, and Alan C. Bovik
Laboratory for Vision Systems, University of Texas, Austin, TX 78712-1084
dave@vision.ece.utexas.edu

Abstract

We present a new theory of multidimensional AM-FM image modeling and derive algorithms for extracting AM-FM sub-image information from digital images. In contrast to Fourier components, AM-FM image functions admit arbitrarily varying amplitude and phase modulations. Thus, they are inherently capable of efficiently capturing essential nonstationary image structures. Often, such nonstationarities contribute significantly to visual perception and interpretation.

We describe a practical approach for computing AM-FM image representations using nonlinear demodulation operators. A Gabor filterbank isolates components locally, and optimal filters based on a statistical state-space component model are used to track image multi-components across the filterbank channel responses. We present dramatic examples where the essential structure of natural images is successfully recovered from their computed AM-FM representations.

1. Introduction

The discrete Fourier transform models an $N \times N$ image as the sum of N^2 sinusoidal components, each with constant amplitude and frequency. For nonstationary analysis, however, it is of interest to consider an image $t(\mathbf{m})$ to be a sum of AM-FM functions

$$t(\mathbf{m}) = \sum_{i=1}^{K} t_i(\mathbf{m}) = \sum_{i=1}^{K} a_i(\mathbf{m}) \exp[j\varphi_i(\mathbf{m})], \qquad (1)$$

where $\mathbf{m} = [m\ n]^T$, $a_i : \mathbb{Z}^2 \to [0, \infty)$, $\varphi_i : \mathbb{Z}^2 \to \mathbb{R}$, and $t, t_i : \mathbb{Z}^2 \to \mathbb{C}$. In (1), it is *assumed* that $t(\mathbf{m})$, $t_i(\mathbf{m})$, $a_i(\mathbf{m})$, and $\varphi_i(\mathbf{m})$ are the samples of corresponding continuous-domain functions $t(\mathbf{x})$, $t_i(\mathbf{x})$, $a_i(\mathbf{x})$, and $\varphi_i(\mathbf{x})$ taken with respect to a unity sampling interval.

A real-valued image $s(\mathbf{x})$ may be analyzed against the model (1) using the *analytic image* $t(\mathbf{x}) = s(\mathbf{x}) + j\mathcal{H}[s(\mathbf{x})]$ [2], where

$$\begin{aligned} \mathcal{H}[s(\mathbf{x})] &= \frac{1}{\pi} \int_{\mathbb{R}} s(\mathbf{x} - \tau \mathbf{e}_x) \frac{d\tau}{\tau} \\ &= s(\mathbf{x}) * \frac{\delta(\mathbf{x}^T \mathbf{e}_y)}{\pi \mathbf{x}^T \mathbf{e}_x}. \end{aligned} \qquad (2)$$

In (2), $\mathbf{e}_x$ and $\mathbf{e}_y$ are horizontal and vertical unit vectors, respectively, and the integral takes its Cauchy principal value. Then, $s(\mathbf{x}) = \mathrm{Re}\,[t(\mathbf{x})]$, the Fourier transform $T(u, v) = \mathcal{F}[t(\mathbf{x})]$ is zero in the left half of the $\mathbf{\Omega} = [u\ v]^T$ frequency plane, and

$$\mathcal{F}\{\mathcal{H}[s(\mathbf{x})]\} = -j\,\mathrm{sgn}(u) S(u, v). \qquad (3)$$

Each component $t_i(\mathbf{m})$ in the sum (1) is an AM-FM image function admitting modulating functions $a_i(\mathbf{m})$ and $\nabla\varphi_i(\mathbf{m})$ that are permitted to vary arbitrarily across the image domain. Hence, each AM-FM component is inherently capable of capturing significant nonstationary image structure, and many images can be represented using *substantially* fewer than N^2 AM-FM components.

In this paper, we compute multi-component representations (1) admitting only smoothly varying, or *locally coherent* components [1]. Such components tend to each embody physically meaningful, smooth, nonstationary or evolutionary structure within an image. Thus, these representations find great utility in nonstationary image analysis. Furthermore, we expect that locally coherent AM-FM image representations will find significant application in future image and video coding schemes; a related 1D approach achieved compression ratios of 43:1 for sampled musical instrument signals [3].

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.

2. Demodulation

In computing representations of the form (1), it is necessary to isolate the multiple image components from one another to facilitate estimation of each component's individual modulating functions. We use a multiband bank of Gabor filters for this purpose. The filters need not isolate components on a global scale; rather, it is required that each filter response be dominated by at most one component at each pixel.

Let $g_p(\mathbf{m})$ be the unit-pulse response of the Gabor filter for filterbank channel p, and let $G_p(\boldsymbol{\omega})$ be the filter frequency response. Let $y_p(\mathbf{m}) = t(\mathbf{m}) * g_p(\mathbf{m})$ and let $\mathbf{e}_k$ represent $\mathbf{e}_x$ or $\mathbf{e}_y$. Then, for the component that dominates $y_p(\mathbf{m})$ at pixel $\mathbf{m}$, each component of $\nabla\varphi_i(\mathbf{m})$ of may be estimated using

$$\mathbf{e}_k^T \nabla\widehat{\varphi}_i(\mathbf{m}) = \arcsin\left[\frac{y_p(\mathbf{m}+\mathbf{e}_k) - y_p(\mathbf{m}-\mathbf{e}_k)}{2jy_p(\mathbf{m})}\right] \quad (4)$$

and

$$\mathbf{e}_k^T \nabla\widehat{\varphi}_i(\mathbf{m}) = \arccos\left[\frac{y_p(\mathbf{m}+\mathbf{e}_k) + y_p(\mathbf{m}-\mathbf{e}_k)}{2y_p(\mathbf{m})}\right] \quad (5)$$

Each of (4), (5) places the estimated frequency vector components to within π radians. Together, they place $\nabla\widehat{\varphi}_i(\mathbf{m})$ to within 2π radians. Once $\nabla\widehat{\varphi}_i(\mathbf{m})$ has been computed, $a_i(\mathbf{m})$ may be estimated by

$$\widehat{a}_i(\mathbf{m}) = \left|\frac{y_p(\mathbf{m})}{G_p[\nabla\widehat{\varphi}_i(\mathbf{m})]}\right|. \quad (6)$$

The approximate algorithm (4) - (6) is based on a discrete *quasi-eigenfunction approximation.* The approximation errors are tightly bounded by theoretical results which guarantee that these errors will generally be small or negligible if the components $t_i(\mathbf{m})$ are each locally coherent [1].

3. Filterbank

We isolate components in the model (1) on a spatio-spectrally localized basis using optimally conjointly localized Gabor filters in a polar wavelet-like frequency tesselation. The continuous-domain unity L^2-norm channel filter $g_p(\mathbf{x})$ with radial center frequency $r_p = |\boldsymbol{\Omega}_p|$ and orientation $\theta_p = \arg\boldsymbol{\Omega}_p$ is

$$g_p(\mathbf{x}) = \frac{1}{\sigma_p\sqrt{2\pi}} \exp\left[-\frac{1}{4\sigma_p^2}\mathbf{x}^T\mathbf{x}\right] \exp\left[j\boldsymbol{\Omega}_p^T\mathbf{x}\right].$$

The frequency response $G_p(\boldsymbol{\Omega})$ is Gaussian, the filter η-peak radial octave bandwidth is $B = \log_2[r_2/r_1]$, where $r_1 = r_p - \sqrt{-\ln\eta}/\sigma_p$ and $r_2 = r_p + \sqrt{-\ln\eta}/\sigma_p$, and the η-peak orientation bandwidth is $\Theta = 2\arctan\sqrt{\gamma}$, where $\gamma = (2^B-1)^2/(2^B+1)^2$. We arrange the filter center frequencies along rays in quadrants I and IV of the frequency plane such that the η-peak contours of any four adjacent filters intersect in a single point. The radial center frequency of the first filter on each ray is r_0, and the center frequencies of the remaining filters on the ray follow a geometric series with common ratio R. The angular spacing between rays is $\Lambda = 2\arcsin\left[(4R)^{-\frac{1}{2}}\left\{(R^2+1)(\gamma-1)+2R(\gamma+1)\right\}^{\frac{1}{2}}\right]$.

To capture the low-frequency information in the image, we incorporate a baseband channel into the filterbank. The impulse response of this filter is

$$g_b(\mathbf{x}) = \frac{1}{\sigma_b\sqrt{2\pi}} \exp\left[-\frac{1}{4\sigma_b^2}\mathbf{x}^T\mathbf{x}\right]. \quad (7)$$

The filter space constant σ_b is is designed so that the η-peak contours of $G_b(\boldsymbol{\Omega})$ and of the first filters on any two adjacent rays of the filterbank all intersect in a single point. Thus,

$$\sigma_b = \frac{\sqrt{-\ln\eta}}{r_0\cos\frac{\Lambda}{2} - \sqrt{\frac{-\ln\eta}{\sigma_0^2} - r_0^2\sin^2\frac{\Lambda}{2}}}, \quad (8)$$

where $\sigma_0 = \sqrt{-\ln\eta}/(2\pi r_0\sqrt{\gamma})$.

The filterbank used in the examples of Section 6 had parameters $r_0 = 9.6$ cycles per image, $R = 1.8$, $B = 1$ octave, and $\eta = 1/2$, giving $\gamma = 1/9$, $\Theta \approx 38.9°$, and $\Lambda \approx 20.6°$. This filterbank has 40 frequency modulated channels and one baseband channel.

4. Component Model

Each channel of the multiband Gabor filterbank delivers estimated modulating functions $\nabla\widehat{\varphi}_i(\mathbf{m})$ and $\widehat{a}_i(\mathbf{m})$ at each pixel in the image. To compute the representation (1), we must determine how many components are present and which filterbank channel should be used to estimate the modulating functions of each component at each pixel.

We extract estimated modulating functions for a low-frequency component from the filterbank baseband channel at each pixel. For the remaining components $t_i(\mathbf{m})$, we formulate a statistical state-space component model in terms of the modulating functions $\nabla\varphi_i(\mathbf{m})$ and $a_i(\mathbf{m})$. Using this model, we design Kalman filters for tracking $\nabla\varphi_i(\mathbf{m})$ and $a_i(\mathbf{m})$ across the filterbank channel responses. The baseband channel estimates are not considered by the Kalman filters.

In formulating the state-space model, we begin by *ordering* the image pixels according to a path function

$\mathcal{P}$ which maps each pixel to a scalar lattice point $k \in \mathbb{N}$. Pixels are mapped sequentially by traversing the first row of the image from left to right, the second row from right to left, and so on. Under the mapping $\mathcal{P}$, the modulating functions of $t_i(\mathbf{m})$ may be written as $\nabla\varphi_i(k)$ and $a_i(k)$. Then, $a_i(k)$ and each component of $\nabla\varphi_i(k)$ may be expanded in a first-order Taylor series about a lattice point k. The derivatives of each modulating function (taken with respect to arc length along $\mathcal{P}$) may also be expanded in zeroth-order Taylor series about k. Writing these Taylor series together leads to a sixth-order canonical state-space model for $t_i(\mathbf{m})$, as described in [2].

The system modes of the state-space model corresponding to $a_i(k)$ and to each component of $\nabla\varphi_i(k)$ may be decoupled to yield three independent second-order systems. The system describing $a_i(k)$ is

$$\begin{bmatrix} a_i(k+1) \\ a_i'(k+1) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_i(k) \\ a_i'(k) \end{bmatrix} + \begin{bmatrix} u_a(k) \\ \nu_a(k) \end{bmatrix} \quad (9)$$

$$a_i(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} a_i(k) \\ u_a(k) \end{bmatrix}, \quad (10)$$

where $a_i'(k)$ is the derivative of $a_i(k)$ and where $u_a(k)$ and $\nu_a(k)$ are uncorrelated stochastic processes [1, 2]. Analogous systems describe the components of $\nabla\varphi_i(k)$.

The estimates (4) - (6) delivered by filterbank channel p are related to the system output (10) by

$$\widehat{a}_i(k) = a_i(k) + n_a(k), \quad (11)$$

where $n_a(k)$ is a stochastic process modeling the errors inherent in (4) - (6). The design of a Kalman filter for tracking and estimating $a_i(k)$ follows from (9) - (11), and is given in [2], along with filters for tracking and estimating each component of $\nabla\varphi_i(k)$.

At each point k along $\mathcal{P}$, the Kalman filters deliver estimates of the quantities $a_i(k)$ and $\nabla\varphi_i(k)$ for each tracked component. They also *predict* what values these quantities will assume at point $k+1$. Observations $\widehat{a}_i(k+1)$ and $\nabla\widehat{\varphi}_i(k+1)$ of component $t_i(\mathbf{m})$ are then taken from the filterbank channel with center frequency closest to the component's predicted frequency. After all existing component tracks have been updated at point k, the filterbank is scanned for channels delivering frequency estimates (4), (5) which do not associate with any existing track. New tracks are initiated using the observations delivered by such channels.

5. Postfilters

The Kalman filters described in Section 4 work well for images that contain only a few components which are well separated in frequency. However, many natural images contain phase discontinuities arising from, *e.g.*, occlusions, surface defects, or surface discontinuities. The modulation function $\nabla\varphi_i(\mathbf{m})$ of an image component $t_i(\mathbf{m})$ can generally contain *unbounded* excursions of either sign near such discontinuities.

If the Kalman filters are permitted to follow such frequency excursions in an image comprising several components, then the track corresponding to a particular component $t_i(\mathbf{m})$ containing a wideband frequency excursion will typically take observations from a filterbank channel that is dominated by another tracked component $t_q(\mathbf{m})$. When this occurs, the tracks corresponding to these components tend to become coincident and continue updating from the same filterbank channels indefinitely.

Therefore, in computing representations for multipartite natural images, it is necessary to postfilter the estimates (4) - (6) to ameliorate the effects of wideband frequency excursions arising from image phase discontinuities. We use low-pass Gaussian postfilters, since their envelopes and bandwidths may be simply related to the channel filters. With $\mathbf{x} = [x\ y]^T$, the impulse response of the unity L^1-norm postfilter for smoothing the estimated modulating functions delivered by filterbank channel p is

$$p_p(\mathbf{x}) = \frac{1}{4\pi\kappa_\theta\kappa_\Omega\sigma_p^2} \exp\left[-\frac{1}{4\sigma_p^2}\left(\frac{\zeta^2}{\kappa_\theta^2} + \frac{\xi^2}{\kappa_\Omega^2}\right)\right], \quad (12)$$

where σ_p is the space constant of channel filter $g_p(\mathbf{x})$, κ_θ and κ_Ω are *postfilter space constant scaling factors*,

$$\begin{bmatrix} \zeta \\ \xi \end{bmatrix} = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (13)$$

$\vartheta = \theta_p - \frac{\pi}{2}$, and θ_p is the orientation of channel filter $g_p(\mathbf{x})$. Thus, the orientation bandwidth of the filter (12) is governed by the scaling factor κ_θ and the magnitude frequency bandwidth is governed by κ_Ω. If $\kappa_\theta = \kappa_\Omega = 1.0$, then the linear bandwidth of $p_p(\mathbf{x})$ is identical to that of $g_p(\mathbf{x})$. Larger postfilter space constant scaling factors yield a filter with a narrower bandwidth, which performs more smoothing. Taking κ_θ small facilitates the movement of tracked components between channels which differ in orientation, while taking κ_Ω small facilitates the movement of tracked components between channels which differ in magnitude frequency. With the postfilters, initial frequency estimates are computed using (4) and (5). These estimates are then postfiltered, and the postfiltered frequency estimates are used in the amplitude algorithm (6). The amplitude estimates themselves are then postfiltered. For the examples of Section 6, the postfilter space constant scaling factors were $\kappa_\theta = 3.0$ and $\kappa_\Omega = 3.25$.

6. Examples

The image *Reptile* is shown in Fig. 1(a). The low-pass component shown in Fig. 1(c) was reconstructed from the estimated modulating functions of the filter-bank baseband channel. This component embodies smooth shading and contrast variations in the image. The Kalman filters detected five additional components in this image. Reconstructions of these components are shown in Fig. 1(d) - (h), and each capture elements of the hexagonal texture structure. The components of Fig. 1(d) and (f) are each dominated by different orientations in different image regions. The amplitude modulations in the components of Fig. 1(e) and (h) are complementary: the component of Fig. 1(e) has lower contrast on the left and higher contrast on the right. This is reversed in the component of Fig. 1(h). Finally, the component of Fig. 1(g) is of relatively lower frequency and may be interpreted as an *AM-FM harmonic* of the component in Fig. 1(f). A reconstruction of the image from all six computed components appears in Fig. 1(b). It is of remarkable quality for such a small number of components. The computed representation has clearly succeeded in capturing the essential structure of the image.

Fig. 1(i) shows the image *Burlap*, while the image *Raffia* is shown in Fig. 1(m). For each of these images, the Kalman filters identified eight components. A reconstruction of the *Burlap* image from all eight tracked components and the baseband component is given in Fig. 1(j). Reconstruction of the *Raffia* image from its computed nine-component representation appears in Fig. 1(n). These reconstructions agree remarkably with the originals, and the essential structure of both images has been captured by the computed representations using *only nine* components.

7. Future Work

While truly amazing, the reconstructed images of Fig. 1(b), (j), and (n) are not perfect. There is noticeable high frequency information missing from each reconstructed image. Such structure tends to be supported on small, irregularly shaped image regions. For example, the prominent vertical edge separating the left one-fourth of the *Reptile* image from the right three-quarters of the image is supported only over a few columns. Thus, it is extremely difficult for the Kalman filters described in Section 4 to maintain tracking of this structure while traversing back and forth across image rows using the path $\mathcal{P}$. Nevertheless, elements of the edge *are* manifest in nonstationary frequency modulations in the components of Fig. 1(d), (f), and (g), as well as in nonstationary amplitude modulations in the components depicted in Fig. 1(c), (e), and (h). Thus, low frequency elements of the edge are clearly visible in the six-component reconstruction of Fig. 1(b). A similar effect occurs in the hexagonal patterns visible in the upper left-hand portion of the image and elsewhere. While the essential elements of these patterns appear in the reconstructed image, there is noticeable high frequency information missing.

In addition to images containing regionally supported structure, the computational approach presented in this paper also has difficulty in handling images which contain multiple objects or multiple regions exhibiting disparate texture characteristics. An alternative approach called the *channelized components paradigm* involves computing one AM-FM component from the estimated modulating functions delivered by each channel of the multiband Gabor filterbank. There are two main disadvantages to this more simplistic approach. First, many of the channelized components tend to be nearly zero over most of the image. Second, various individual features of the image structure tend to be manifest in multiple channelized components. Thus, channelized components representations are inefficient.

Nevertheless, we typically obtain excellent reconstructions from channelized components AM-FM image representations. For example, Fig. 1(l) shows a channelized components reconstruction of the *Reptile* image. Channelized components reconstructions of the *Burlap* and *Raffia* images are shown in Fig. 1(k) and (o), respectively. Finally, Fig. 1(p) shows a channelized components reconstruction of the image *Celebrity*.

Future research in computed multi-component AM-FM image representations will involve the development of improved Kalman filtering strategies to obtain high quality reconstructions such as those shown in Fig. 1(k), (l), (o), and (p) using only a few locally coherent computed AM-FM components.

References

[1] J. P. Havlicek, D. S. Harding, and A. C. Bovik. Discrete quasi-eigenfunction approximation for AM-FM image analysis. In *Proc. IEEE Int'l. Conf. Image Proc.*, pages 633–636, Lausanne, Switzerland, September 16-19, 1996.

[2] J. P. Havlicek, D. S. Harding, and A. C. Bovik. The mutli-component AM-FM image representation. *IEEE Trans. Image Proc.*, 5(6):1094–1100, June 1996.

[3] J. A. Moorer. Signal processing aspects of computer music: A survey. *Proc. IEEE*, 65(8):1108–1137, August 1977.

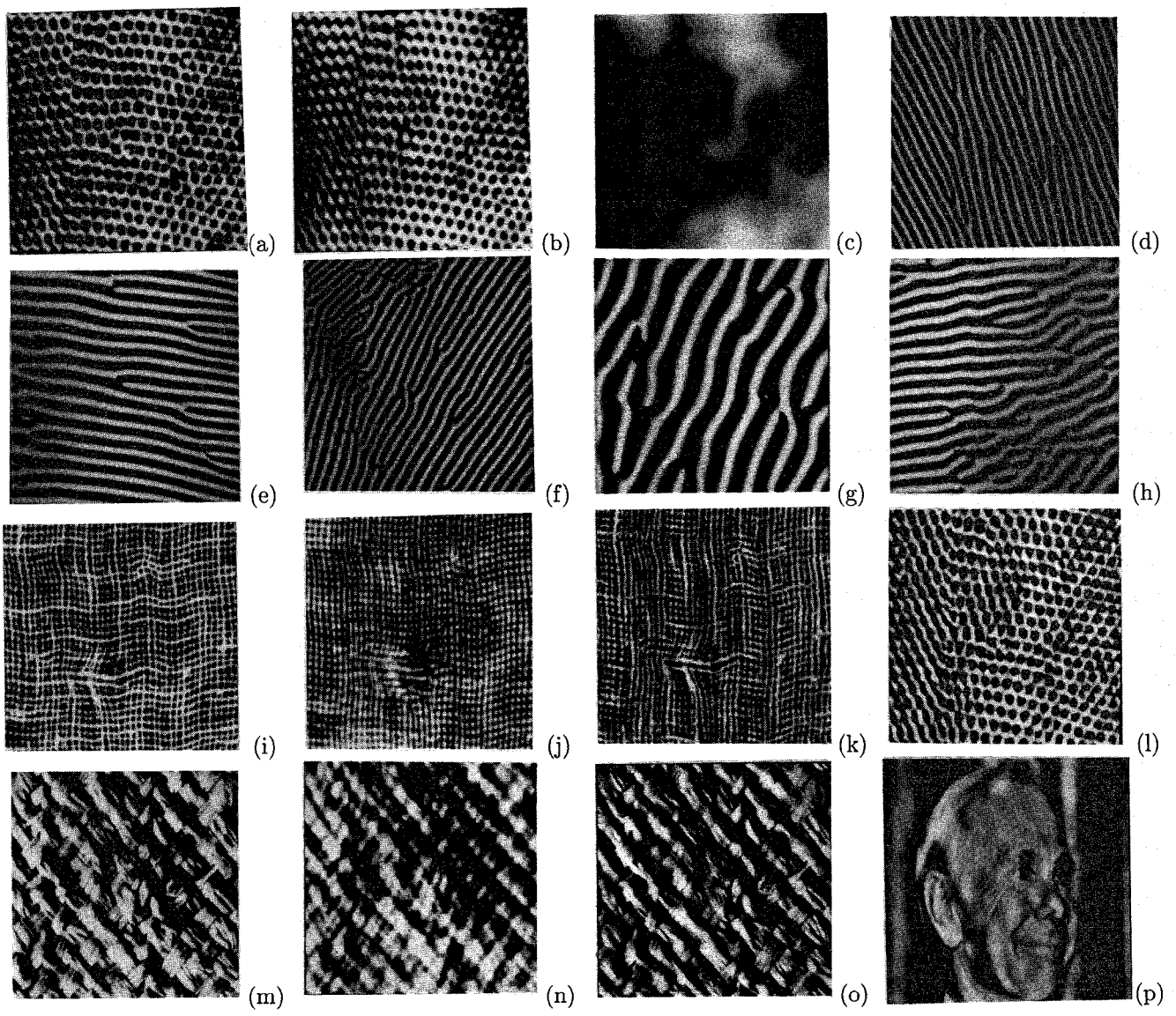

Figure 1. Computed AM-FM image representations. (a) Reptile image. (b) Reconstruction from six computed AM-FM components. (c) - (h) Reconstructed components of Reptile image. (i) Burlap image. (j) Reconstruction from nine computed AM-FM components. (k) Reconstruction from channelized AM-FM components. (l) Reconstruction of Reptile image from channelized AM-FM components. (m) Raffia image. (n) Reconstruction from nine computed AM-FM components. (o) Reconstruction from channelized AM-FM components. (p) Reconstruction of Celebrity image from channelized AM-FM components.