

COHERENT TEXTURE DECOMPOSITION USING AM-FM MODEL

Chuong T. Nguyen and Joseph P. Havlicek
School of Electrical and Computer Engineering
University of Oklahoma, Norman, OK, USA
chuong@ou.edu, joebob@ou.edu

Abstract—We introduce a novel decomposition algorithm capable of extracting locally coherent and visually meaningful texture components from images. The algorithm estimates texture dominant orientation for each coherent component and iteratively extracts it from the image based on a new quantitative coherency measure formulated in the modulation domain. The original image is perfectly reconstructed from extracted components.

Keywords—image decomposition; AM-FM image model

I. INTRODUCTION

Separating intrinsic component from a complex signal has been one of the fundamental tasks in the signal and image processing arena [1]–[5]. Intrinsic components not only enable an in-depth understanding of complex signals but also provides unique features for higher level applications, *e.g.*, recognition and classification. The signal decomposition, however, is particular challenging because it seeks to solve for more than one solutions given one input signal, *viz.*, ill-posed problem [4], [5]. For example, in the cocktail party problem, one attempts to isolate voice of individuals from a conversation involving multiple people and background noise; or in image analysis where the objective is to decompose an image of multiple texture sources into *visually meaningful* components.

Early attempts employed multiresolution time-frequency techniques to represent the signal as a sum of localized parts. Examples of these techniques include Gabor filterbank by Daugman [6], the steerable pyramid by Simoncelli and Freeman [7], and wavelets [8]. While these techniques have been applied successfully in many applications, they frequently produce components that fail to correspond with human visual perception. This limitation is a consequence of lacking specific perceptually motivated models for components.

Recently, Meyer [9] pioneered a two component image decomposition using nonlinear partial differential equation (PDE) approach. In this approach, an image is broken into a sum of two parts: a cartoon and a texture. The cartoon depicts a the structure of the image and is modeled by a bounded variational function, whereas the texture part is modeled by certain energy norms. These two components

This work was supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant W911NF-08-1-0293.

are computed simultaneously by a total variation minimization framework [3], [4]. Similar to Meyer’s cartoon and texture decomposition idea, Starck, *et al.* [5] introduced an image decomposition where components are extracted from two sparse dictionaries optimized for cartoon and texture. Evangelopoulos and Maragos [10] also adopted this two component image model and formulated the decomposition in the modulation domain. Even though the two component image decomposition model delivers meaningful results, it is restricted to two components and does not produce locally coherent components.

In contrast to the cartoon and texture model, Bovik, *et al.* [1] proposed a multicomponent image model where they represented each component with slowly varying amplitude modulation (AM) and frequency modulation (FM) functions, thereby explicitly computing AM-FM image models in the modulation domain. A Kalman filtering framework was developed in [11] to track texture multicomponents spatially across the channels of a Gabor filterbank and extract them. While this approach did not prove sufficiently robust to enable reliable analysis of general images, it should be noted that the 1-D AM-FM signal model was applied successfully in [12] to track multicomponent amplitude and frequency modulations temporally in human speech.

In this paper, we perform image decomposition using the modulation domain image model (AM-FM model) in [1], [13]. We propose an iterative algorithm for decomposing a texture image into homogeneous components. We estimate the dominant orientations for the coherent components and introduce a quantitative coherency measure to isolate them from each other. The coherent components are then iteratively extracted by a greedy matching algorithm similar to matching pursuit [2]. The extracted components are locally coherent and agree with human perception. In addition, they can perfectly reconstruct the original image.

II. AM-FM TEXTURE DECOMPOSITION

The AM-FM image model represents an image $t(\mathbf{x})$ as a finite sum of K AM-FM components [1], [11]

$$t(\mathbf{x}) = \sum_{k=1}^K t_k(\mathbf{x}) = \sum_{k=1}^K a_k(\mathbf{x}) \cos[\varphi_k(\mathbf{x})], \quad (1)$$

where the AM functions $a_k(\mathbf{x})$ and FM functions $\nabla\varphi_k(\mathbf{x})$ are smoothly varying and $\mathbf{x} \in \mathbb{R}^2$. In an image, we interpret $a_k(\mathbf{x})$ as local contrast, $|\nabla\varphi_k(\mathbf{x})|$ as texture spacing or granularity, and $\arg\nabla\varphi_k(\mathbf{x})$ as local texture orientation.

Let $m, n \in \mathbb{N}$ and let $\mathbf{I}(m, n)$ contain the samples of the continuous image $t(\mathbf{x})$ in (1). Let $\mathbf{I}_k(m, n)$ contain the samples of component $t_k(\mathbf{x})$ in (1). Then the component $\mathbf{I}_k(m, n)$ admits a modulation domain representation $\Gamma_k = [A_k(m, n) \ R_k(m, n) \ \theta_k(m, n)]$ which may be computed with the spline-based demodulation framework given in [14], where

$$\begin{aligned} R_k(m, n) &= |\nabla\varphi_k(m, n)|, \\ \theta_k(m, n) &= \arg\nabla\varphi_k(m, n), \end{aligned} \quad (2)$$

and $A_k(m, n)$ is the magnitude of the complex image whose real part is the image \mathbf{I}_k and imaginary part is the partial Hilbert transform of the real part. We construct the modulation domain representation of a K component image $\mathbf{I}(m, n)$ by stacking individual components as

$$\mathbf{\Gamma} = [\Gamma_1 \ \Gamma_2 \ \dots \ \Gamma_K]^T. \quad (3)$$

The vector $\mathbf{\Gamma}$ will be used as a dictionary for component extractions.

We compute the dominant FM $\vartheta(m, n)$ of the image $\mathbf{I}(m, n)$ by selecting the maximum of all R_k on a pixel-by-pixel basis. We estimate the orientation of the coherent component $\mathbf{I}_k(m, n)$ by constructing an eight-bin histogram from the dominant orientation $\vartheta(m, n)$. The orientation angle ρ is obtained by searching for the histogram's peak. The orientation map of coherent component $\mathbf{I}_k(m, n)$ is then given as

$$\Theta_k(m, n) = \begin{cases} \vartheta(m, n), & |\vartheta_k(m, n) - \rho| \leq \pi/16, \\ \rho, & \text{otherwise.} \end{cases} \quad (4)$$

We illustrate the process of computing orientation map $\Theta_k(m, n)$ for each component in Fig. 1. The original cloth image is shown in Fig. 1(a) while its dominant FM field is shown in Fig. 1(b). The dominant FM field in Fig. 1(b) contains arrows pointing to multiple orientations. The proposed orientation estimation algorithm is able to iteratively decompose it into two dominant orientation maps in Fig. 1(c) and Fig. 1(d) that are each highly coherent, both locally and globally.

The dominant orientation (4) is used as unique feature to match against the overcomplete dictionary $\mathbf{\Gamma}$ (3). The texture component is computed as the linear combination

$$D(m, n) = \sum_{j=1}^K \alpha_j(m, n) A_j(m, n) \cos[\varphi_j(m, n)], \quad (5)$$

where the coherency measure $\alpha_j(m, n)$ is defined by

$$\alpha_j(m, n) = \begin{cases} 1, & \delta_j(m, n) \leq \frac{\pi}{16}, \\ \exp\left(-\frac{\delta_j(m, n) - \pi/16}{\pi/16}\right), & \text{otherwise} \end{cases} \quad (6)$$

and $\delta_j(m, n) = |\theta_j(m, n) - \Theta_k(m, n)|$. Intuitively, the coherency measure $\alpha_j(m, n)$ indicates the *closeness* of the orientation field (4) to the modulation functions in the dictionary $\mathbf{\Gamma}$. The threshold bandwidth $\pi/16$ in (6) was chosen for agreement with the eight orientations present per level in the adapted steerable pyramid.

Since the algorithm performs image decomposition recursively, it allows perfect reconstruction of the original image from the decomposed components. The coherent components are extracted from the image iteratively until the energy of extracted component is less than 0.1% the energy of the original image or the energy of residual signal is no more than 1% compared to that of the original. Consequently, residual component does not contain any significant textures of the original image and it can be interpreted as the DC component.

III. RESULTS AND DISCUSSION

We tested the algorithm against a variety of Brodatz textures. For each extracted component, we computed the modulation domain feature dictionary $\mathbf{\Gamma}$ as in (3). We then estimated the dominant texture orientation Θ as in (4). The extracted component D was then found by matching the dominant orientation Θ against the dictionary $\mathbf{\Gamma}$ as described in (5) and (6). The results are illustrated in Fig. 2. The first row shows the original test images. The second and third rows illustrate the first and the second components respectively. The fourth row shows the DC component of the decomposition. The fifth row depicts the reconstructed image from the first two components and the DC component. The gray scale levels in these images are contrast stretched together and directly comparable.

The original *woven brass* image is shown in Fig. 2(a). It can be interpreted as a two component image with perpendicular diagonal stripes. The algorithm produces five components; three components are shown Fig. 2(b)-(d). The first two components in Fig. 2(b)-(c) are locally coherent and coincide with human perception. Fig. 2(e) shows the reconstruction from Fig. 2(b), (c), and (d). The reconstructed image from the first two components captures most of the visual impact of the original image with 89% energy.

Another perceptually motivated decomposition example is illustrated for the *raffia* image in Fig. 2(f). The algorithm is able to identify oriented textural lines in the raffia image and decompose them into two dominant components shown in Fig. 2(g)-(h). The reconstruction image using the first three components contains 80% of the original energy and is depicted in Fig. 2(j). Vertical and horizontal components of the *loose burlap* image in Fig. 2(k) are presented in Fig. 2(l)-(m). The DC component is shown in Fig. 2(n), while the three component reconstruction image is depicted in Fig. 2(o). The reconstructed image from the first three components contains 90% of the original energy.

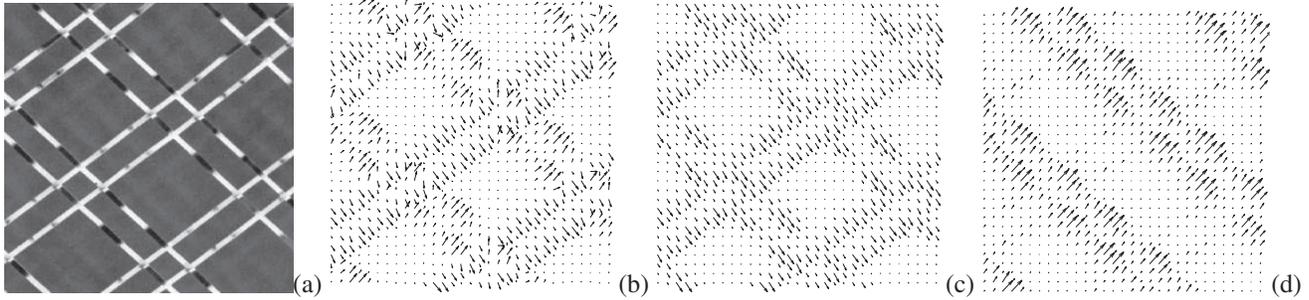


Figure 1. (a) Original woven brass image. (b) Dominant FM field of woven brass. (c) First dominant orientation. (d) Second dominant orientation.

Although the extract components in Fig. 2(q)-(r) of the tree image are locally coherent, they do not agree with human perception which interprets the image with one circular component. Such limitation can be explained by the range restriction imposed on the angular bandwidth threshold of the coherency measure in (6). The angular bandwidth was set to $\pi/16$, which is not wide enough to handle circular textural patterns.

IV. CONCLUSIONS

In this paper, we proposed an iterative texture decomposition algorithm capable of extracting locally coherent and meaningful components from textural images. We introduced a new quantitative coherency measure in the modulation domain for image components. The algorithm provides a perfect reconstruction decomposition. Reconstructed image from first few components is able to capture faithfully the visual impact of the original image. The effectiveness of the proposed algorithm is demonstrated with a variety of well-known Brodatz textures. The decomposed image components are visually motivated and their interpretations coincide with human perception.

The proposed algorithm, however, is limited to texture images with components having limited orientation bandwidth. The future work will be focused on fine-tuning the coherency measure in (6) to accommodate circularly symmetric textural patterns and apply the algorithm to nature images.

REFERENCES

- [1] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 1, pp. 55–73, Jan. 1990.
- [2] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Proc.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [3] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Sci. Comput.*, vol. 19, pp. 553–577, Oct. 2003.
- [4] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition-modeling, algorithms, and parameter selection," *Int'l. Journal Comput. Vision*, vol. 67, no. 1, pp. 111–136, Apr. 2006.
- [5] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Imag. Proc.*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.
- [6] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am. A*, vol. 2, no. 7, pp. 1160–1169, Jul. 1985.
- [7] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transform," *IEEE Trans. Info. Theory*, vol. 38, no. 2, pp. 587–607, March. 1992.
- [8] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal., Machine Intel.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [9] M. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: The fifteenth Dean Jacqueline B. Lewis memorial lectures*. AMS, 2001.
- [10] G. Evangelopoulos and M. Maragos, "Texture modulation-constrained image decomposition," in *Proc. Int'l. Conf. Imag. Proc.*, San Diego, CA, Oct. 12–15, 2008, pp. 793–796.
- [11] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "The multi-component AM-FM image representation," *IEEE Trans. Image Proc.*, vol. 5, no. 6, pp. 1094–1100, Jun. 1996.
- [12] S. Lu and P. C. Doerschuk, "Nonlinear modeling and processing of speech based on sums of AM-FM formant models," *IEEE Trans. Signal Proc.*, vol. 44, no. 4, pp. 773–782, Apr. 1996.
- [13] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models," *IEEE Trans. Image Proc.*, vol. 9, no. 2, pp. 227–242, Feb. 2000.
- [14] R. A. Sivley and J. P. Havlicek, "Perfect reconstruction AM-FM image models," in *Proc. IEEE Int'l. Conf. Image Proc.*, Atlanta, GA, Oct. 8–11, 2006, pp. 2125–2128.

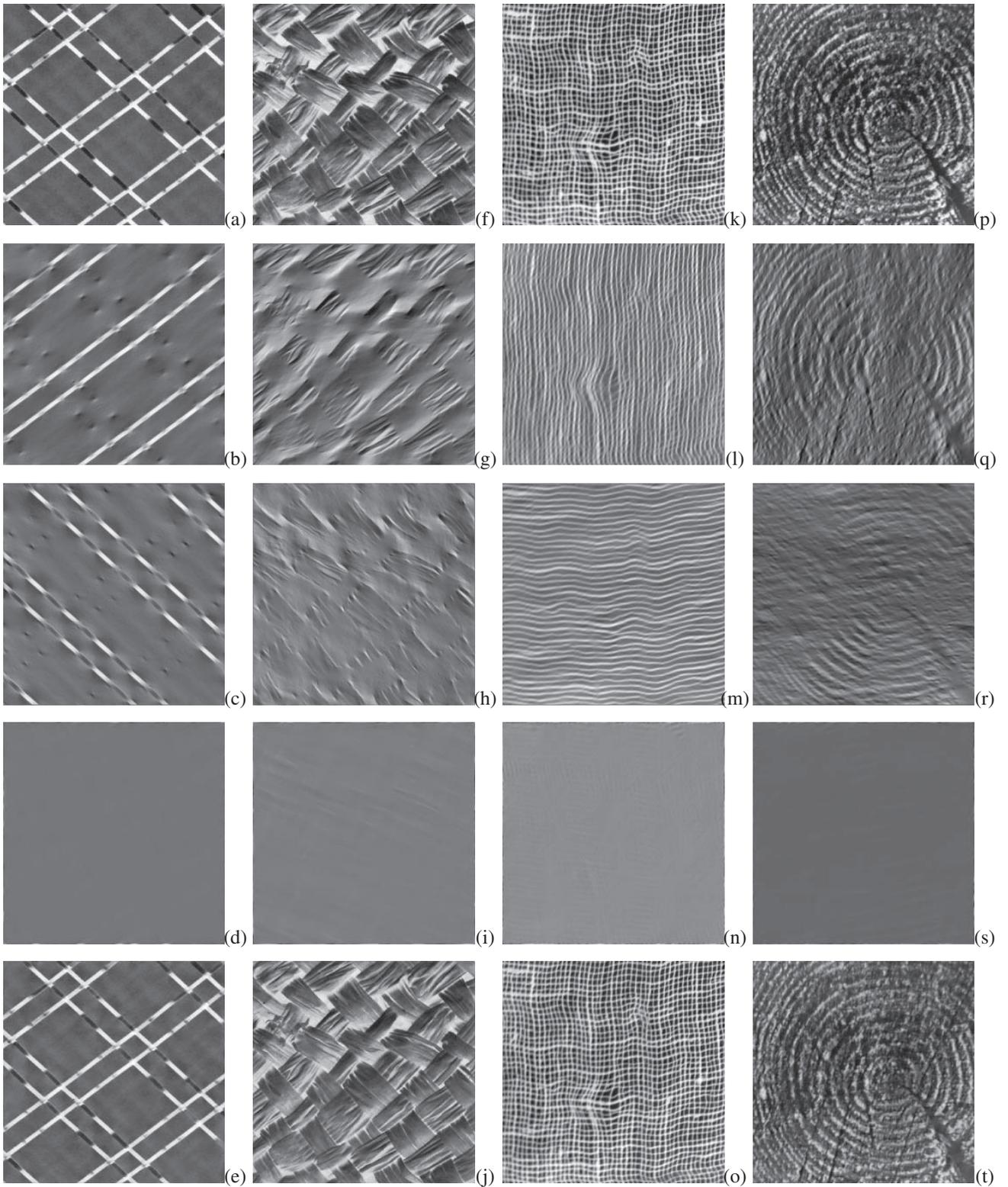


Figure 2. Texture Decomposition Examples. (a) Original woven brass image. (b) First component of woven brass. (c) Second component of woven brass. (d) DC component of woven brass. (e) Reconstruction of (b), (c), and (d). (f) Original raffia image. (g) First component of raffia. (h) Second component of raffia. (i) DC component of raffia (j) Reconstruction of (g), (h) and (i). (k) Original burlap. (l) First component of burlap. (m) Second component of burlap. (n) DC component of burlap. (o) Reconstruction of (l), (m) and (n). (p) Original tree. (q) First component of tree. (r) Second component of tree. (s) DC component of tree. (t) Reconstruction of (q), (r) and (s).