

MODULATION DOMAIN TEXTURE DECOMPOSITION

Chuong T. Nguyen and Joseph P. Havlicek

School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK, USA

ABSTRACT

We introduce a novel texture analysis algorithm capable of extracting *visually meaningful* and locally coherent sub-textural components from images. The algorithm operates in the modulation domain where texture is represented by locally coherent amplitude and frequency modulation functions. The texture components are iteratively extracted based on a new quantitative coherency measure. The effectiveness of the algorithm is demonstrated on several well-known Brodatz textures.

Index Terms— AM-FM image model, texture segmentation, texture analysis, local coherency

1. INTRODUCTION

Decomposing a complicated signal into perceptually meaningful components is an important problem that has received increasing attention recently [1–4]. Well-known examples include the cocktail party speech separation problem and image restoration from multiple sub-image sources. In this paper, we are interested in the image decomposition problem where a texture image is broken into multiple *visually meaningful* components, *e.g.*, simple and locally coherent constituents. Unfortunately, such decomposition is an ill-posed inverse problem [3, 4]. Starck, *et al.* [4] illustrated that a K component image of N pixels will require $N \times K$ unknowns to be solved. Therefore, prior knowledge of the signal components should be incorporated, *e.g.*, signal statistics, image models, and sparsity.

Early approaches used multiresolution techniques to describe the image as a sum of localized parts. Daugman [5] proposed a human visual system (HVS) inspired Gabor filterbank to decompose an image into smooth and localized components. The computed components are band-pass and orientation selective. Alternatively, Simoncelli and Freeman [6] introduced the steerable pyramid for image analysis. The steerable pyramid decomposes an image into multiple scales and multiple orientations. In the past two decades, wavelets have been used extensively in denoising and compression applications. Wavelets offer good time-frequency localization and a compact representation [7]. While these techniques are essential in many image processing applications, they lack specific models for coherent texture components. Consequently, the decomposed components obtained by these analysis techniques frequently fail to correspond well with human visual perception of the image.

Recently, Meyer [8] pioneered a nonlinear partial differential equation (PDE) approach to image decomposition. In this formulation, an image is broken into a sum of two parts: a cartoon part and a texture part. The cartoon describes a homogeneous region with sharp boundaries, and is modeled by a bounded variational function. The texture part is modeled by certain energy norms. Both

of these components are computed simultaneously by a total variation minimization framework [2, 3]. Similar to Meyer's cartoon and texture decomposition idea, Starck, *et al.* [4] formulated an image decomposition using a combination of basis pursuit denoising and total-variation regularization. They used two optimized and sparse dictionaries, one for the cartoon and one for the texture, to extract image components. Even though the two component image decomposition model delivers meaningful results, this approach does not generate locally coherent components.

In contrast to the two component image model approach, cartoon and texture, Havlicek, *et al.* [9, 10] attempted to decompose images into sums of multiple locally narrow-band components. 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 [9] 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 extended Kalman filtering was applied successfully in [11] to track multicomponent amplitude and frequency modulations temporally in human speech. The spatially adaptive Kalman filters of [9] were replaced by a static global decomposition into components based on the filterbank structure in [10], which led to a robust and readily computable multicomponent image model. However, such decomposition precluded the possibility of perfect reconstruction and produced components that were less strongly connected to human visual perception. Evangelopoulos and Maragos [12] also used the modulation domain model for image decomposition, but their approach was limited to the two component decomposition model, cartoon and texture.

In this paper, we propose a novel iterative algorithm for decomposing a texture image into homogeneous textural patches that are locally coherent and visually meaningful. The components are iteratively extracted by a greedy algorithm that is similar to matching pursuit [1] and utilizes a new quantitative modulation domain coherency measure. The experimental results show that the extracted components are locally coherent and agree well with human perception. We discuss the modulation domain image model in Section 2. The texture decomposition algorithm is given in Section 3. Results and discussion appear in Section 4, while conclusions are given in Section 5.

2. MODULATION DOMAIN IMAGE MODEL

The AM-FM image model represents an image $t(\mathbf{x})$ as a finite sum of K AM-FM components [9, 13]

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

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.

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. The model has been successfully used in a variety of image processing applications [14].

Prior to computing the modulation functions in (1) for each component, the image $t(\mathbf{x})$ is usually decomposed into smooth and localized channel images $t_k(\mathbf{x})$ by filtering through a filterbank, *e.g.*, Gabor filter bank, wavelets, or steerable pyramid. In this paper, we use an adapted overcomplete steerable pyramid consisting of five levels and eight orientations per level. This adapted implementation is based on the original steerable pyramid introduced by Simoncelli and Freeman [6].

The AM and FM functions in (1) were estimated using the Teager-Kaiser energy operator in [15]. An alternative approach, *e.g.*, analytic image demodulation, computes these modulation functions as

$$\begin{aligned} a_k(\mathbf{x}) &= |z_k(\mathbf{x})|, \\ \nabla\varphi_k(\mathbf{x}) &= \operatorname{Re} \left[\frac{\nabla z_k(\mathbf{x})}{j z_k(\mathbf{x})} \right], \end{aligned} \quad (2)$$

where $z_k(\mathbf{x})$ is a complex image given by

$$z_k(\mathbf{x}) = t_k(\mathbf{x}) + j\mathcal{H}\{t_k(\mathbf{x})\} \quad (3)$$

and $\mathcal{H}\{\cdot\}$ is the partial adjusted Hilbert transform [10]. Given the continuous component $t_k(\mathbf{x})$, the algorithm (2) is exact. For a discrete image, it can still be used to compute discrete modulation functions by utilizing the spline-based framework introduced in [16].

3. MODULATION DOMAIN TEXTURE DECOMPOSITION

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 $\mathbf{I}_k(m, n)$ admits a modulation domain representation $\Gamma_k = [A_k(m, n) \ R_k(m, n) \ \theta_k(m, n)]$ which may be computed using (2) with the spline-based demodulation framework given in [16], where $R_k(m, n) = |\nabla\varphi_k(m, n)|$ and $\theta_k(m, n) = \arg\nabla\varphi_k(m, n)$. The overall K component image $\mathbf{I}(m, n)$ is described in the modulation domain by the multicomponent representation $\Gamma = [\Gamma_1 \ \Gamma_2 \ \dots \ \Gamma_K]^T$ obtained by concatenating the representations of the individual components. The vector Γ is then used as a dictionary for matching orientations as the dominant texture components are extracted.

However, the FM functions $\nabla\varphi_k(m, n)$ in (2) typically contain undesirable artifacts at frequencies oriented along the direction of action of the partial Hilbert transform. These artifacts result primarily from the discrete approximation of the continuous partial Hilbert transform filter. To ameliorate this problem, we actually compute two FM functions for $\mathbf{I}_k(m, n)$ using two different partial Hilbert transforms acting in orthogonal directions: a transform \mathcal{H}_1 with horizontal action and a transform \mathcal{H}_2 with vertical action. Intuitively, we think of this as steering the action of the transform away from the important frequency content of the component, so that at least one of the two FM functions will be relatively free from Hilbert transform artifacts. The FM function with greater frequency magnitude is selected on a pixel-by-pixel basis along with the corresponding AM for inclusion in the representation Γ_k .

Let A_k^1, R_k^1 , and θ_k^1 be the AM, FM magnitude, and FM orientation functions computed for component $I_k(m, n)$ using the

Table 1. AM-FM Texture Decomposition Algorithm

```

L ← original image  $\mathbf{I}(m, n)$ 
domOrient ←  $\emptyset$ 
while(true)
  1. Compute feature vector  $\Gamma$  for image L using (2).
  2. Estimate the dominant orientation  $\Theta(m, n)$  using (4).
     if ( $\Theta \in$  domOrient) break;
  3. Extract texture component  $C_d$  using (6) and (7).
  4.  $L \leftarrow L - C_d$ 
  5. domOrient ← [domOrient  $\Theta$ ]
end while

```

\mathcal{H}_1 , and let A_k^2, R_k^2 , and θ_k^2 be defined similarly for the \mathcal{H}_2 transform. For $q = 1, 2$ we compute AM-weighted FM functions

$$\begin{aligned} \mathcal{R}^q(m, n) &= \sum_{k=1}^K A_k^q(m, n) R_k^q(m, n) \\ \Theta^q(m, n) &= \sum_{k=1}^K A_k^q(m, n) \theta_k^q(m, n). \end{aligned}$$

At each pixel, we compute the dominant FM orientation given by

$$\Theta(m, n) = \Theta^{l(m, n)}(m, n), \quad (4)$$

where

$$l(m, n) = \arg \max_{q \in \{1, 2\}} \{\mathcal{R}_q(m, n)\}. \quad (5)$$

The dominant orientation (4) is matched against the overcomplete dictionary Γ to assign a weight to the AM and FM functions in each Γ_k on a pixel-by-pixel basis. The dominant texture component is then defined by the linear combination

$$C_d(m, n) = \sum_{k=1}^K \alpha_k(m, n) A_k(m, n) \cos[\varphi_k(m, n)], \quad (6)$$

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

$$\alpha_k(m, n) = \begin{cases} 1, & \delta_k(m, n) \leq \frac{\pi}{12}, \\ 1 - \left\{ \frac{3}{\pi} [\delta_k(m, n) - \frac{\pi}{12}] \right\}^2, & \frac{\pi}{12} < \delta_k(m, n) \leq \frac{\pi}{6}, \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

and $\delta_k(m, n) = |\theta_k(m, n) - \Theta(m, n)|$. Intuitively, the coherency measure α_k works in local spatial neighborhoods to group together and blend components $\mathbf{I}_k(m, n)$ having FM orientations that are *close* to the dominant orientation (4) at each pixel. The blended sum (6) then constitutes a coherent, textural component that is extracted from the image and the process is repeated iteratively as indicated in Table 1 until all coherent dominant orientations have been extracted from the image. The threshold bandwidth $\pi/6$ in (7) was chosen for agreement with the eight orientations present per level in the adapted steerable pyramid.

The proposed algorithm can be interpreted as a frequency-based feature extraction technique. It is, however, different from other traditional techniques such as Fourier and orientation-selective Gabor filtering. Our algorithm is capable of finding the dominant texture orientation of a component and extract it by matching its orientation against an overcomplete feature dictionary. Other frequency-based techniques rely on an energy-based analysis step to estimate texture orientations, and subsequently require one to design filters with appropriate bandwidths and orientations for the feature extraction process. Both of these estimation steps are sensitive to errors if the image structures and orientations are not known a priori.

4. RESULTS AND DISCUSSION

We tested the algorithm described in Table 1 against a variety of Brodatz textures. For each extracted component, we computed the modulation domain feature dictionary Γ . The dominant texture orientation Θ was estimated as in (4). The extracted component $C_d(m, n)$ was then found by matching the dominant orientation $\Theta(m, n)$ against the dictionary Γ as described in (6) and (7). The results are illustrated in Fig. 1. The left column contains original images. The second and third columns show the first and second dominant components and the last column shows the residual after extraction of the first two dominant components. Images in the second, third, and fourth columns of each row are contrast stretched together and directly comparable in terms of gray scales.

The original burlap image is shown in Fig. 1(a). It can be interpreted as a two component image with horizontal and vertical stripes. The algorithm produces two dominant components shown in Fig. 1(b)-(c). These components are locally coherent and coincide with human perception. The lack of organized texture in the residual image in Fig. 1(d) demonstrates that the horizontal and vertical stripes of the original image are successfully extracted.

A more complex decomposition example of reptile skin texture is shown in Fig. 1(e). The two dominant texture components are shown in Fig. 1(f)-(g). The algorithm is capable of decomposing the original image into three locally coherent image components. It is interesting to observe that these components are not easily identified from the original image. Similar perceptually motivated results can be seen for the straw image in Fig. 1(i). Straws having vertical orientations are successfully grouped into the first coherent component in Fig. 1(j). The second extracted component in Fig. 1(k) depicts remaining textures that are orthogonal to the first.

Fig. 1(m) shows a composite image which contains a wood grain texture on the left and a uniform grainy texture on the right. The decomposed texture components are shown in Fig. 1(n)-(o). The first component in Fig. 1(n) is able to capture the wood grain texture part of the original image, while the second component in Fig. 1(o) extracts the remaining texture. This example suggests that better decomposition results can be achieved if the original image is segmented into homogeneous regions prior to applying this algorithm.

The proposed algorithm, however, is not effective for circularly symmetric texture images like the tree image in Fig. 1(q). Although the extract components in Fig. 1(r) and Fig. 1(s) 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 (7). The angular bandwidth was set to $\pi/6$, which is not wide enough to handle circular textural patterns.

5. CONCLUSION

In this paper, we proposed an iterative texture analysis 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 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 (7) to

accommodate circularly symmetric textural patterns and apply the algorithm to nature images.

6. REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] M. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: The fifteenth Dean Jacqueline B. Lewis memorial lectures*, AMS, 2001.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] G. Evangelopoulos and M. Maragos, "Texture modulation-constrained image decomposition," in *Proc. Int'l. Conf. Image Proc.*, San Diego, CA, Oct. 12-15, 2008, pp. 793–796.
- [13] 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.
- [14] C. T. Nguyen and J. P. Havlicek, "AM-FM image filters," in *Proc. IEEE Int'l. Conf. Image Proc.*, San Diego, CA, Oct. 12-15, 2008, pp. 789–792.
- [15] P. Maragos and A. C. Bovik, "Image demodulation using multidimensional energy separation," *J. Opt. Soc. Amer. A*, vol. 12, no. 9, pp. 1867–1876, Sep. 1995.
- [16] 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.

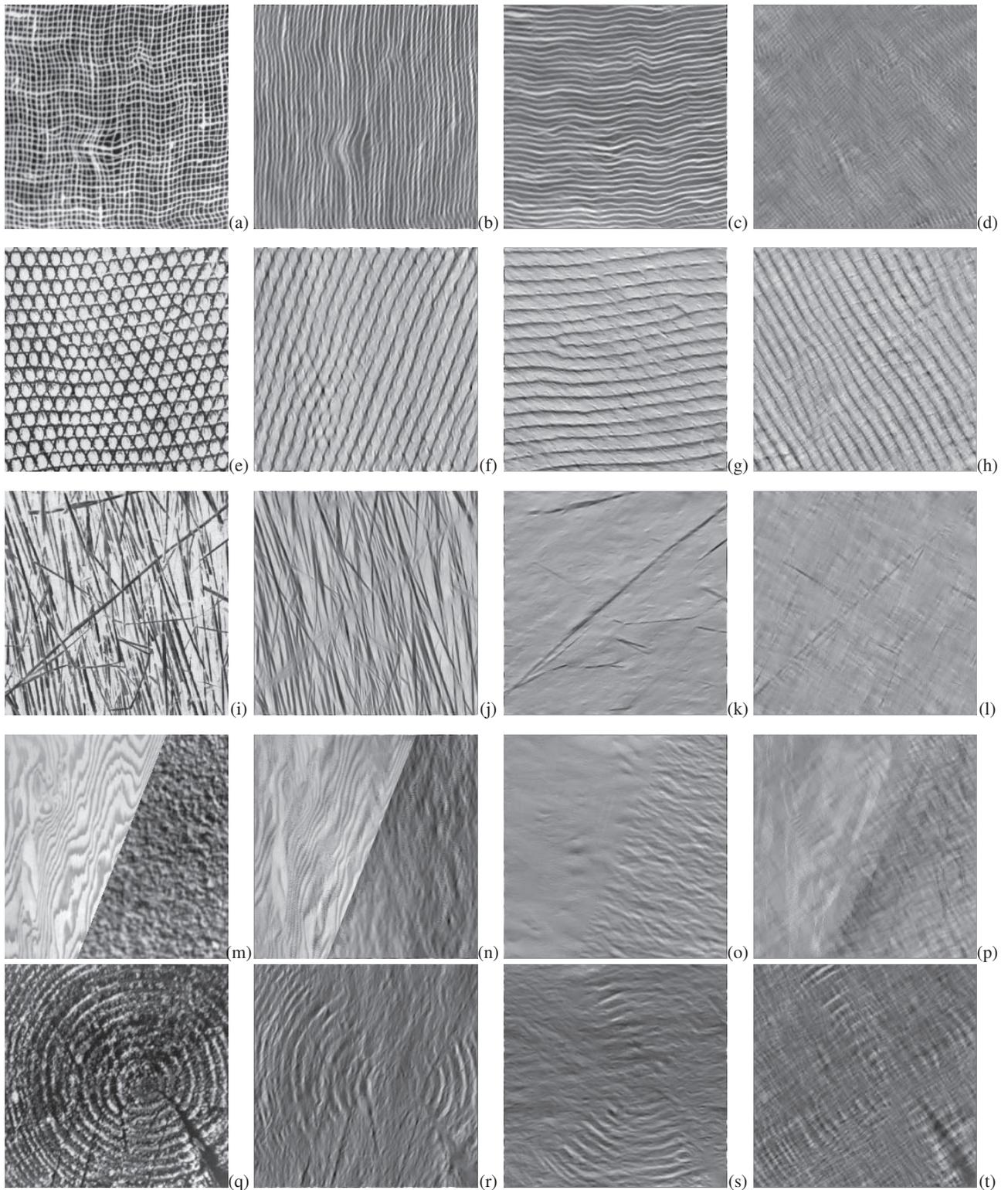


Fig. 1. Examples. (a) Original burlap image. (b) First component of burlap. (c) Second component of burlap. (d) Residual of burlap. (e) Original reptile skin image. (f) First component of reptile skin. (g) Second component of reptile skin. (h) Residual of reptile skin. (i) Original straw image. (j) First component of straw. (k) Second component of straw. (l) Residual of straw. (m) Original wood/paper image. (n) First component of wood/paper. (o) Second component of wood/paper. (p) Residual of wood/paper. (q) Original tree image. (r) First component of tree. (s) Second component of tree. (t) Residual of tree.