Texture Analysis and Segmentation Using Modulation Features, Generative Models, and Weighted Curve Evolution

Iasonas Kokkinos, *Member*, *IEEE*, Georgios Evangelopoulos, *Member*, *IEEE*, and Petros Maragos, *Fellow*, *IEEE*

Abstract—In this work, we approach the analysis and segmentation of natural textured images by combining ideas from image analysis and probabilistic modeling. We rely on AM-FM texture models and, specifically, on the Dominant Component Analysis (DCA) paradigm for feature extraction. This method provides a low-dimensional, dense, and smooth descriptor, capturing the essential aspects of texture, namely, scale, orientation, and contrast. Our contributions are at three levels of the texture analysis and segmentation problems: First, at the feature extraction stage, we propose a *Regularized Demodulation Algorithm* that provides more robust texture features and we explore the merits of *modifying the channel selection criterion* of DCA. Second, we propose a probabilistic interpretation of DCA and Gabor filtering in general, in terms of *Local Generative Models*. Extending this point of view to edge detection facilitates the estimation of *posterior probabilities for the edge and texture classes*. Third, we propose the *Weighted Curve Evolution* scheme that enhances curve evolution-based segmentation methods by allowing for the locally adaptive combination of heterogeneous cues. Our segmentation results are evaluated on the Berkeley Segmentation Benchmark and compare favorably to current state-of-the-art methods.

Index Terms—Texture analysis, image segmentation, AM-FM models, demodulation, generative models, curve evolution, cue combination.

1 Introduction

Texture is ubiquitous in natural images and constitutes a powerful cue for a variety of image analysis and computer vision applications like segmentation, shape from texture, and image retrieval. The advances of the last two decades in image analysis and biological and computer vision have deepened our understanding of this field, yet it remains open and challenging.

The problem of texture analysis has been addressed using primarily feature and model-based methods; feature-based methods [2], [22], [30], [44], [47], [57] analyze texture using an informative description that lends itself more easily to subsequent tasks, typically using linear filterbanks as front-end systems. Members of the second category, like Markov Random Fields (MRFs) [8], [56], use tractable models for texture patterns and formulate texture analysis as a parameter estimation task; the gap between these two approaches has been bridged in [17], [56], yielding a powerful yet intricate common framework. A different path has pursued the use of textons [23]; an operational definition of textons as cluster centers in a filter response space is advocated in [31], [37], while, in [16], [17], a texton dictionary is proposed as a medium for the optimal representation of images.

These are powerful models for texture analysis, but their appropriateness for unsupervised texture segmentation is limited in some respects. In conjunction with both boundary-based [31], [32], [37] and region-based [30], [31], [44], [50], [57] approaches, the high dimensionality of filterbank features can lead to poor segmentations and requires dimensionality reduction, which is a problem in itself. MRF-based approaches suffer from a computational aspect since their fitting is coupled with segmentation, resulting in a time-consuming iterative procedure. Texton-based approaches fit naturally with pairwise clustering techniques [31], [50], where the proximity between two pixels is estimated by comparing the distributions of texton indexes in their neighborhoods. However, such descriptors cannot be used by variational and generative segmentation methods alike [1], [30], [44], [52], [57] that rely on having smooth features within homogeneous regions.

Our approach builds on the class of Amplitude-Modulation-Frequency-Modulation (AM-FM) image models [18], [19], [34] and, specifically, on the Dominant Component Analysis (DCA) method [20]. In short, DCA represents texture locally in terms of a single AM-FM signal, whose parameters are estimated and used as a texture descriptor. This yields a feature set that encompasses information about texture contrast, scale, and orientation while lending itself naturally to tasks like density estimation used in image segmentation.

In our work, whose preliminary versions have been presented in [11], [27], [28], [29], we pursue the construction

- *I. Kokkinos is with the Department of Applied Mathematics, Ecole Centrale Paris, Grande Voie des Vignes, 92295 Chatenay-Malabry, France and INRIA-Saclay. E-mail: iasonas.kokkinos@ecp.fr.*
- *G. Evangelopoulos and P. Maragos are with the School of Electrical and Computer Engineering, National Technical University of Athens, Zografou Campus, 15773, Athens, Greece. E-mail: {gevag, maragos}@cs.ntua.gr.*

Manuscript received 15 May 2006; revised 30 Jan. 2007; accepted 15 Nov. 2007; published online 30 Jan. 2008.
Recommended for acceptance by H. Shum.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0380-0506.
Digital Object Identifier no. 10.1109/TPAMI.2008.33.

of a concise texture analysis and segmentation system for generic natural images by extending the potential of the DCA method. Specifically, our contributions to texture analysis, feature interpretation, and texture segmentation are listed as follows:

1. *Feature Extraction*. In Section 2.2, a regularized algorithm for demodulation is introduced which avoids discrete image differentiations using combinations of Gabor filtering and the 2D Teager-Kaiser energy operator [34], [35]. The potential of alternative criteria for channel selection based on the 2D operator is explored in Section 2.3, yielding features that are more appropriate for segmentation.
2. *Probabilistic Analysis*. A probabilistic formulation of the AM-FM channel selection procedure is presented in Section 3 by modeling observations in terms of sinusoids and introducing locality in the likelihood expressions. This facilitates the interpretation of Gabor filtering in terms of model fitting, which is a formulation we also use in Section 3.3 to phrase edge detection in common terms with texture analysis. This lays the ground for the probabilistic discrimination between edges, textured areas, and smooth areas, which is a practically important problem for image segmentation.
3. *Image Segmentation*. In Section 4, we present an unsupervised segmentation scheme based on DCA features that uses curve evolution implemented with level set methods. Using our probabilistic analysis results, we propose a method for the combination of heterogeneous cues that enhances the original Region Competition-Geodesic Active Region (GAR) evolution rule [44], [57]. Specifically, we introduce the Weighted Curve Evolution method that incorporates the posterior probabilities of the texture and edge classes in the evolution law. We report systematic experiments on the Berkeley Benchmark, where consistent improvements in performance are attained when compared to simpler or different segmentation methods.

Since our contributions span different levels of the overall analysis and segmentation system, each section is written in a modular manner, with introductory subsections on prior work and necessary background information.

2 AM-FM Texture Models

Locally narrow-band signals can model a variety of textured images like patterns formed by surface deformations, orientation-diffusion biological markings, and man-made objects exhibiting periodic structure, like those in Fig. 1. Modulation, or AM-FM models, have been successfully applied to speech signal analysis [4], [35] and are ideally suited for the description of such image signals [3], [34]. Modeling signals in terms of nonstationary sinusoids,

$$f(x,y) = a(x,y)\cos(\phi(x,y)), \tag{1}$$

AM-FM models locally capture image contrast in terms of the *amplitude modulating signal* $a(x,y)$ and image structure (scale and orientation) in terms of the *instantaneous frequency vector*:

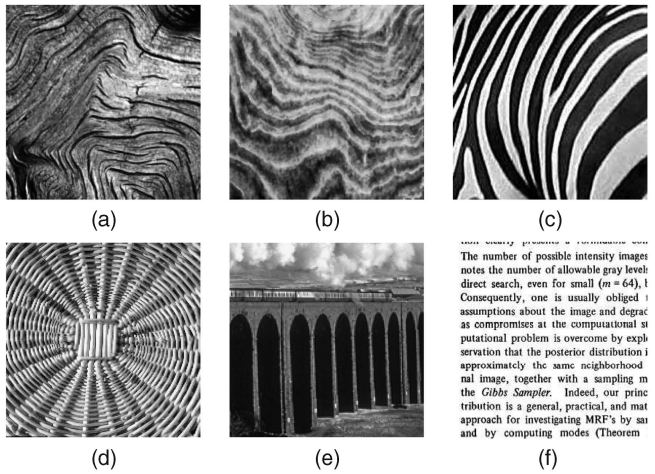


Fig. 1. Textures of the locally narrowband type. (a) Results of evolutionary processes. (b) Surface deformations. (c) Biological patterns. (d)-(f) Periodic man-structured objects.

$$\vec{\omega}(x,y) = \nabla\phi(x,y) = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}\right)(x,y). \tag{2}$$

Even though many natural textures can be modeled in terms of a monocomponent AM-FM signal, images with 2D structure containing patterns like corners, crosses, and junctions necessitate that more than one component be simultaneously present in the local image spectrum. The multicomponent AM-FM model [19], [20] models an image I as the superposition of locally narrow-band sinusoidal components $f_k(x,y)$ corrupted by a white Gaussian noise (WGN) field $w(x,y)$:

$$I(x,y) = \sum_{k=1}^{K} \underbrace{a_k(x,y)\cos(\phi_k(x,y))}_{f_k(x,y)} + w(x,y). \tag{3}$$

The fundamental problem of image demodulation aims at estimating, for each of the K components, the instantaneous amplitudes $a_k(x,y)$ and frequencies $\vec{\omega}_k(x,y) = \nabla\phi_k(x,y)$.

The decomposition of an image in terms of this expression is an ill-posed problem due to the existence of an infinity of modulating signal pairs and component superpositions satisfying (3). Even if a separation of I in narrowband components $f_k(x,y)$ is known in advance, unavoidable modeling errors of any demodulation algorithm, the presence of noise, interference from neighbor spectral components, and discretization of the signal derivatives are possible sources of error in component estimation. Robustness in the AM-FM demodulation problem can be achieved by considering the following problems:

- P1. Reduction of the error in modeling each narrowband component $f_k(x,y)$ by a 2D AM-FM signal while maintaining smoothness in the estimated modulation signals.
- P2. Suppression of noise.
- P3. Suppression of neighbor spectral components while estimating one component.
- P4. Regularization of derivatives.

Simultaneously achieving all of the above goals is a complex optimization task which remains an unsolved problem. In the following sections, well-established solutions

to problems P1-P3 are presented, followed, in Section 2.2, by a novel algorithm that jointly considers all problems. In Section 2.3, the DCA method is presented, together with a modified channel selection criterion that yields better localized features.

2.1 AM-FM Demodulation

2.1.1 Energy Operators and Demodulation

At the heart of problem P1 lies the fact that there are an infinite number of combinations that satisfy (1) for a given f. An efficient scheme for the demodulation of the narrow-band components into smooth modulating functions is provided by the multidimensional *Energy Separation Algorithm* (ESA) [34], which is based on a generalization to higher dimensions of the 1D Teager-Kaiser energy operator [35]:

$$\Psi(f)(x,y) \triangleq \|\nabla f(x,y)\|^2 - f(x,y)\nabla^2 f(x,y). \quad (4)$$

Now, let f be a 2D spatial AM-FM signal as in (1). Under realistic assumptions [34], applying Ψ to f yields the energy product of the squared instantaneous amplitude and frequency magnitude:

$$\Psi[a\cos(\phi)] \approx a^2\|\vec{\omega}\|^2, \quad (5)$$

with an approximation error bounded within a negligible range. This quantity may be interpreted as the component *modulation energy*. Applying Ψ to the partial derivatives $f_x = \partial f/\partial x$, $f_y = \partial f/\partial y$ and combining all energies yields the 2D continuous ESA [34]:

$$\frac{\Psi(f)}{\sqrt{\Psi(f_x)+\Psi(f_y)}} \approx |a(x,y)|, \quad (6)$$

$$\sqrt{\frac{\Psi(f_x)}{\Psi(f)}} \approx |\omega_1(x,y)|, \quad \sqrt{\frac{\Psi(f_y)}{\Psi(f)}} \approx |\omega_2(x,y)|, \quad (7)$$

which can estimate, at each location (x,y), the amplitude envelope and the magnitudes of the instantaneous frequencies of the nonstationary AM-FM signal. The signs of the frequency signals can be implicitly obtained by the signs of the carrier, approximated by the filter central frequencies.

2.1.2 Multiband Gabor Filtering and Demodulation

A simultaneous solution to problems P2 and P3 has been given in [3], [4] using a bank of bandpass filters densely covering the frequency plane. The filterbanks used for this task are typically 2D Gabor filters, favored due to their optimal joint spatial and spectral localization [14], [9]. Apart from component decoupling and robustness to noise, this approach specifies in advance the number and spectral localization of the different components, thereby constraining the decomposition of any given 2D signal to a fixed component configuration. In Fig. 2, we visually show the filterbank used in our experiments, while details are given in Appendix A.

Demodulation via the ESA can be extended to the complex signals derived from convolution with complex Gabor filters; the energy for a complex-valued signal $f(x,y) = a(x,y)\exp(j\phi(x,y))$ is defined as

$$C(f) = \Psi[\mathrm{Re}\{f\}] + \Psi[\mathrm{Im}\{f\}] \quad (8)$$

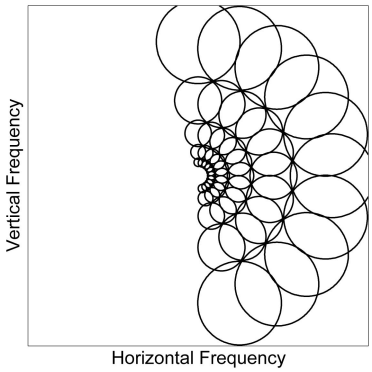


Fig. 2. Filterbank grid on the 2D frequency domain. Contours correspond to half-peak bandwidth magnitude.

and, based on the approximation (5), the operator response is $C[f] \approx 2a^2\|\vec{\omega}\|^2$. The averaging of operator responses results in smoother estimates of the modulating functions. Applying C to $f = I * g$ and its partial derivatives f_x and f_y results in a demodulation scheme where the frequencies are given by (7) and the amplitude is given by a slight modification of (6):

$$|a(x,y)| \approx \frac{C(f)}{\sqrt{2}\sqrt{C(f_x)+C(f_y)}}. \quad (9)$$

Another point is that Gabor filtering imposes a specific decomposition of an arbitrary signal of the form (3) into a sum of narrow-band components, with the frequency content of each component localized around the corresponding Gabor filter's central frequency. However, the frequency content of the actual component may not be centered at the fixed central frequency of the Gabor filter, thereby resulting in a suppressed estimate a_k of its amplitude A_k. This can be compensated for by using the component's estimated instantaneous frequency $\vec{\omega}_k$; specifically, if $G_k(\cdot)$ is the frequency response of the Gabor filter, the approximation

$$A_k = \frac{a_k}{|G_k(\vec{\omega}_k)|} \quad (10)$$

yields an amplitude estimate that is insensitive to deviations from the corresponding filter central frequency [20].

2.2 Regularized Demodulation

A problem that emerges with ESA demodulation is that the signal derivatives can only be approximated using discrete differentiation operations. As a result, the two differential operators entailed in the energy operator responses may furnish inaccurate amplitude and frequency estimates. In what follows, we present a theoretically sound approach to alleviate this problem, introducing a regularized 2D energy operator and a related regularized 2D ESA.

As analyzed in [46] for edge detection, two *regularized solutions* to the derivative estimation problem which minimize the sum of the data approximation error and the energy of the second derivative of the approximating function are 1) spline interpolation and 2) convolution of the image data by a function that can be closely modeled by a Gaussian. In our problem, which deals with narrow-band but not necessarily low-pass signals, the Gaussian filter

TABLE 1
Demodulation Comparisons
between Gabor ESA and Discrete ESA

$\sqrt{<(\hat{A}-A)^2>}$ for Gabor / discrete ESA (bold/plain)				
$\log SNR$	α=0		α=1	
10	$\mathbf{4.2\ 10^{-4}}$	$1.2\ 10^{-2}$	$\mathbf{2.1\ 10^{-2}}$	$3.9\ 10^{-2}$
6	$\mathbf{6.9\ 10^{-4}}$	$1.3\ 10^{-3}$	$\mathbf{2.2\ 10^{-4}}$	$3.9\ 10^{-2}$

$\sqrt{<(\hat{\omega}_x-\omega_x)^2>}$ for Gabor / discrete ESA (bold/plain)				
$\log SNR$	α=0		α=1	
10	$\mathbf{6.4\ 10^{-5}}$	$1.9\ 10^{-2}$	$\mathbf{4.9\ 10^{-3}}$	$2.2\ 10^{-2}$
6	$\mathbf{1.1\ 10^{-4}}$	$1.9\ 10^{-2}$	$\mathbf{4.9\ 10^{-3}}$	$2.3\ 10^{-2}$

response must be modulated by a sine with a carrier equal to the spectral mean location of the signal. This yields a Gabor filter. In [10], the spline and the Gabor regularization of the energy operator and the ESA were compared for 1D signals, yielding a slight superiority of the Gabor ESA.

Motivated by the above, we propose a *2D Gabor ESA* algorithm for simultaneous filtering and demodulation. Let $I(x,y)$ be the continuous image, $g(x,y)$ be the impulse of a real 2D Gabor filter, and $f(x,y)=I(x,y)*g(x,y)$ be its output. Since convolution commutes with differentiation, the continuous 2D energy operator combined with Gabor bandpass filtering becomes

$$\Psi(f)=\Psi(I*g)=\|I*\nabla g\|^2-(I*g)(I*\nabla^2 g). \tag{11}$$

Thus, the differential operators have been replaced by filter derivatives that can be analytically estimated, thereby avoiding discretization errors.

Similarly, for the estimation of the instantaneous amplitude and frequency, the 2D Gabor ESA for demodulating $f=I*g$ consists of the following two steps:

1. Use the Gabor energy operator to compute the instantaneous energies of three image functions, $\Psi(f)$, $\Psi(f_x)$, and $\Psi(f_y)$, where
$$\Psi(f_x)=\|I*\nabla g_x\|^2-(I*g_x)(I*\nabla^2 g_x). \tag{12}$$
2. Use the evaluated energies in the formula of the 2D continuous ESA.

For all three energies, we need seven Gabor differential formulas: g_x, g_y, g_{xx}, g_{yy}, g_{xy}, and $\nabla^2 g_x$, $\nabla^2 g_y$. The Gabor ESA is thus computationally more intensive since it requires more convolutions but adds robustness and improved performance. For efficiency, we use an FFT-based frequency-domain implementation of the Gabor ESA, using the equation

$$\mathcal{F}\left\{\frac{\partial^{k+\ell} g}{\partial x^k \partial y^\ell}\right\}=\mathcal{F}\{g\}(j\omega_x)^k(j\omega_y)^\ell, \tag{13}$$

relating the Fourier transforms $\mathcal{F}\{\cdot\}$ of a signal and its derivatives.

In Table 1, the performance of the discrete ESA is compared to the Gabor-ESA scheme at varying degrees of noise and nonstationarity. Signals of the form

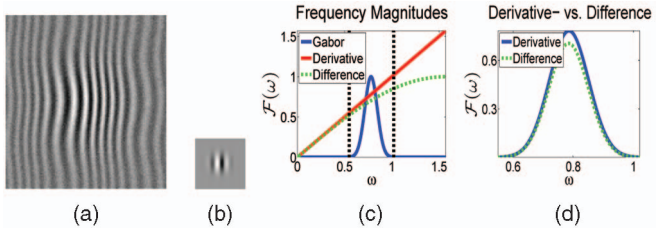


Fig. 3. Regularized demodulation. (a) Representative AM-FM signal of the family (14) obtained for modulation index $\alpha=.5$ and $\log(SNR)=6$. (b) Gabor filter used for demodulation. (c) Fourier transform magnitudes for the filters involved in the alternative demodulation schemes, demonstrating the deviation of the central difference filter Δ_x from the derivative operation $\frac{\partial}{\partial x}$. (d) Deviation of $\Delta_x * g$ from $\frac{\partial g}{\partial x}$ in the frequency domain.

$$f(x,y)=[1+\alpha A(x,y)]\cos(u_c x+v_c y+\alpha\theta(x,y)), \tag{14}$$

$$\theta(x,y)=\frac{1}{4}\left[2\cos\left(\frac{u_c}{30}x\right)+\cos\left(\frac{v_c}{30}y\right)\right], \tag{15}$$

$$A(x,y)=\exp\left(-\frac{x^2+y^2}{10}\right), \tag{16}$$

are used, where u_c and v_c are the central frequencies of the Gabor filter used for demodulation, shown in Fig. 3b. The signal is immersed in white Gaussian noise at various Signal-to-Noise Ratios (SNRs), while the index α is varied to produce different degrees of nonstationarity.

For $\alpha=0$, i.e., a stationary sinusoid, the approximation in (5) becomes exact, so, for Gabor ESA, the only source of error is noise. On the contrary, the differentiation scheme used in the discrete ESA introduces systematic errors, as shown in Fig. 3d, and results in inferior frequency and amplitude estimates. For higher degrees of nonstationarity, Gabor ESA systematically yields better estimates, with the errors being solely due to the noise signal and the approximations of ESA.

2.3 Texture Features

The demodulation procedure furnishes a 3D vector $(A_k,\nabla\phi_k)(x,y)$ for each of the components in (3), so demodulating the filterbank channel outputs yields a $3\times K$-dimensional texture feature vector at each pixel. This multidimensional feature extraction scheme, termed Channelized Component Analysis in [20], provides a rich image representation and can achieve accurate reconstructions of multicomponent signals; however, the high dimensionality of the feature vector may result in poor segmentations.

A compact texture description can be extracted using the *DCA* method [18], [20] that retains the most prominent structure of the texture signal. Assuming that a single narrowband component dominates the filter responses at pixel (x,y), DCA selects pixelwise the channel $i(x,y)$ that is closest to the component, demodulates its output, and uses the resulting AM-FM features for texture representation. The channel $i(x,y)$ is chosen among the K filter responses by maximizing a criterion $\Gamma_k(x,y)$:

$$i(x,y)=\arg\max_{1\le k\le K}\{\Gamma_k(x,y)\}, \tag{17}$$

$$A_{\mathrm{DCA}}(x,y)=A_{i(x,y)}(x,y),\quad \vec{\omega}_{\mathrm{DCA}}(x,y)=\vec{\omega}_{i(x,y)}(x,y). \tag{18}$$

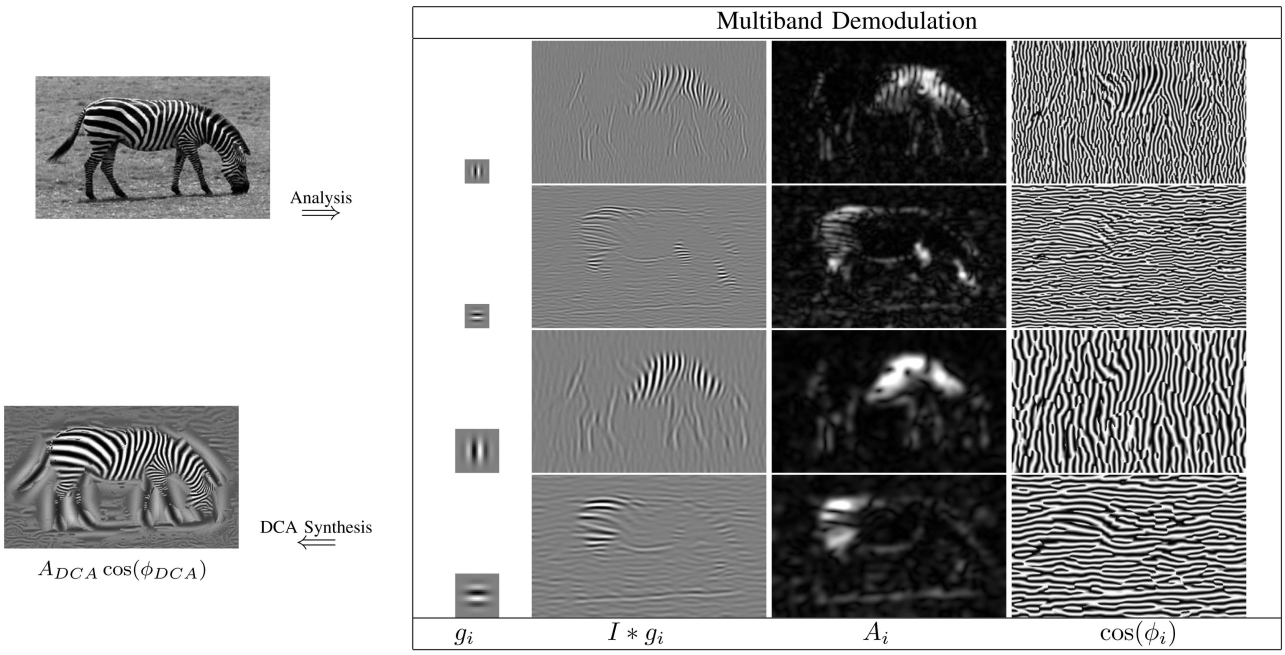


Fig. 4. The Dominant Componnets Analysis method for a locally narrowband signal: A set of bandpass Gabor filters is initially used to isolate and demodulate the individual components of (3). The dominant channel is subsequently chosen at each image location and its AM-FM parameters are used as a local texture descriptor. The principal structure of the textured signal is thus captured by the DCA parameters.

The choice of the dominant channel in the original work on DCA has been based on the maximization of the estimated amplitude envelopes:

$$\Gamma_k(x,y) = |a_k(x,y)|. \tag{19}$$

In Fig. 4, a locally narrowband signal is used to demonstrate the structure-capturing properties of this procedure. A texton dictionary-based method would break the image into pieces indicating which of the textons best match the input signal, yielding a discrete texton-index tessellation of the image, while a filterbank-based feature descriptor would retain all filter responses, even though most offer no information complementary to that of the most active filter. On the other hand, using the DCA method, a single filter is automatically selected and a low-dimensional smoothly varying feature vector is derived from it. Note that, instead of the instantaneous frequency measurements in Fig. 4, we use the phase estimate delivered by the complex Gabor filter since it is better suited for visual display.

The refined frequency and amplitude estimates (7), (10) furnished by the demodulation algorithm thus allow us to transcend from the quantized set of orientations and scales used by the front-end filterbank to a continuous representation.

2.3.1 Energy-Based Dominant Component Analysis (EDCA)

As an alternative to amplitude-based dominant component extraction, termed ADCA henceforth, we have considered an energy channel selection criterion, based on the modulation product (5), leading to the EDCA scheme. Intuitively, if we think of texture signals as produced by physical oscillating sources in different scales and orientations, the selection of the dominant component could be based on the maximum-energy source that accounts for producing the local texture modulations. According to this scheme, modulation features are chosen from the filter output of dominating energy:

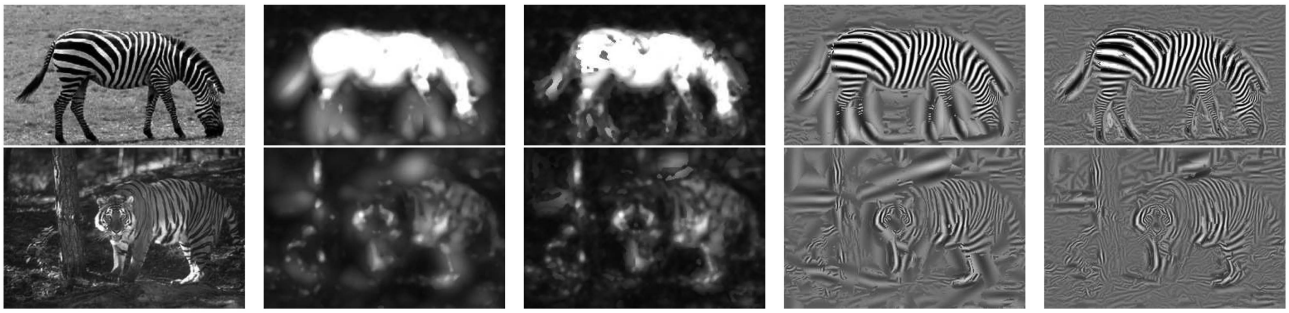
$$\Gamma_k(x,y) = \Psi[(I * g_k)](x,y), \tag{20}$$

where the complex energy operator (8) is used for a complex filter g_k.

Using the modulation energy for DCA results in improved localization in texture and object boundaries: Since the 2D energy operator jointly captures contrast and frequency information in the modulation product (5), the scheme can effectively consider channels with low amplitude (i.e., contrast) variations but high instantaneous frequency magnitude.

To illustrate their differences, in Fig. 5, we compare the features extracted using the original and the alternative energy-based method. Comparing Figs. 5b and 5c, we see that the EDCA measurements are sharper around object boundaries, with improved localization and detail preservation. We observe, for example, that the diffusion effects around the borders of the tiger and the zebra are alleviated using EDCA. The reconstructions delivered by the two schemes reveal the preservation of finer structure in the energy-based scheme; as an indicative example, notice that ADCA interprets the feet of the zebra as a slowly varying horizontal oscillation, while EDCA focuses on the smaller scale structure of the vertical zebra skin pattern.

We note here that the DCA model is designed primarily for 1D features like sinudoidal signals and requires additional AM-FM components to model 0D and 2D features like blobs and crosses, respectively. It would be beneficial to account for such patterns in our front-end system, but we have practically observed that, as seen also in Fig. 5, for images exhibiting such patterns, a perceptually



(a) (b) (c) (d) (e)

Fig. 5. Amplitude versus Energy-based DCA: Comparing the estimated amplitude and the synthesized component using the two alternative channel selection criteria, (19) and (20), we observe that EDCA focuses on more prominent texture variations, combining information about frequency and contrast. Instead, DCA favors large-scale image variations, which are not always perceived as texture. (a) Input image. (b) ADCA amplitude. (c) EDCA amplitude. (d) ADCA synthesis. (e) EDCA synthesis.

meaningful part of the image structure is captured by the DCA features.

3 Local Generative Models for Texture and Edges

In this section, we probabilistically justify the channel selection of DCA, introducing a generative model that accounts for the locality of the decision process. Based on this model, Gabor filtering can be interpreted as parameter estimation. An analogous model is provided for edge detection, allowing us to estimate posterior probabilities for the texture and edge classes, based on Bayes' rule. After a brief introduction of generative models in Section 3.1, in Section 3.2, we present our local generative models and relate them to the DCA method. Section 3.3 extends this approach to edge detection and the discrimination between edges and texture.

3.1 Generative Models

Generative models are capable of reproducing an image or parts of it and probabilistically relate the image observations with the model synthesis. They can thus be used for both modeling and classification, based on Bayes' rule.

Specifically, given K alternative classes $\mathcal{C}_{1,\ldots,K}$ and an observation O, each class uses a low-dimensional parameter set $\mathbf{E}_i$ in its *synthesis equation* $O(x) \simeq I_i(x|\mathbf{E}_i)$ to approximate the observations. Adopting a probabilistic *error model* yields a *likelihood expression* $P_i(O|\mathbf{E}_i)$ for the observations conditioned on the model synthesis and by integrating out the model parameters we obtain the *model evidence*:

$$\begin{aligned} P(O|\mathcal{C}_i) &= \int P(O)|\mathbf{E}_i, \mathcal{C}_i) P(\mathbf{E}_i|\mathcal{C}_i) d\mathbf{E}_i \\ &= \int P_i(O|\mathbf{E}_i) P_i(\mathbf{E}_i) d\mathbf{E}_i, \end{aligned} \tag{21}$$

where $P_i(\mathbf{E}_i)$ is the prior distribution of the model parameters. This integration is typically bypassed, assuming that the ith model evidence is proportional to the maximum value of the integrand $P_i(O|\hat{\mathbf{E}}_i)P_i(\hat{\mathbf{E}}_i)$ attained at the Maximum A Posteriori (MAP) or, if the prior is uniform, the Maximum-Likelihood (ML) parameter estimate $\hat{\mathbf{E}}_i$. Based on this approximation, the Generalized Likelihood Ratio Test (GLRT) [24] assigns the observations to the class with maximal posterior probability, derived from Bayes' rule:

$$P(\mathcal{C}_i|O) = \frac{P(\mathcal{C}_i)P(O|\mathcal{C}_i)}{\sum_{k=1}^{K} P(\mathcal{C}_k)P(O|\mathcal{C}_k)} \simeq \frac{P_i(O|\hat{\mathbf{E}}_i)P_i(\hat{\mathbf{E}}_i)}{\sum_{k=1}^{K} P_k(O|\hat{\mathbf{E}}_k)P_k(\hat{\mathbf{E}}_k)}, \tag{22}$$

where it is assumed that the prior probabilities $P(\mathcal{C}_i)$ for all classes are equal.

For the problems we are interested in, low-level models accounting for generic image variations at the level of patches are needed. Herein, we consider three general signal classes, namely, texture, edges, and smooth areas, with each class using a predetermined parametric synthesis equation.

3.2 Local Generative Models for Texture

Starting with the texture class hypothesis, we develop a link between the DCA method and generative models. For this, we build on the assumption behind DCA, namely, that texture can be locally described in terms of a single narrow-band signal. Intuitively, the Gabor filtering and channel selection stages at the front end roughly estimate the frequency of the signal, using a quantized set of orientations and frequency magnitudes. Here, we establish a formal connection between these two stages and a parametric probabilistic model: We show that the amplitude estimate for a Gabor filter output is proportional to a lower bound on the log likelihood of the observations under a corresponding model hypothesis.

For simplicity, we model the 1D profile of the signal along the feature's orientation, using the frequency magnitude $\omega = |\vec{\omega}|$, while the patch being modeled is considered centered around $x_0 = 0$. Wherever discrete notation is used, the signals are represented as $N \times 1$ column vectors.

3.2.1 Local Modeling of Texture

Our model uses a linear basis that describes the signal as a sinusoid with frequency ω, unknown phase offset ϕ_o, amplitude A, and DC value B:

$$O(x) \simeq I_T(x|\{A, \phi_o, B\}) = A\cos(\omega x + \phi_o) + B \tag{23}$$

$$= \sum_{i=1}^{D_T} E_i B_{T,i}(x), \quad \text{where} \quad D_T = 3, \tag{24}$$

$$E_1 = A\cos(\phi_o), \quad E_2 = -A\sin(\phi_o), \quad E_3 = B, \tag{25}$$

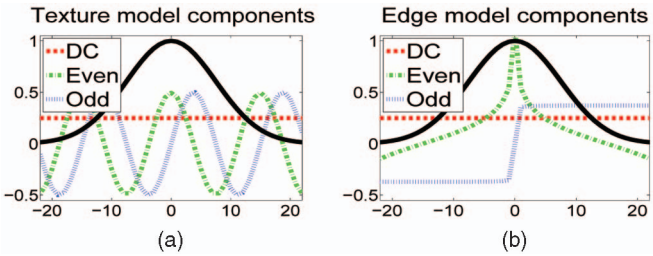


Fig. 6. Basis elements (dashed) and confidence function (solid) for the 1D profiles of the texture and edge classes.

$$B_{T,1}(x) = \cos(\omega x), \quad B_{T,2}(x) = \sin(\omega x), \quad B_{T,3}(x) = 1. \quad (26)$$

The subscript T denotes the texture hypothesis, while we shall refer to the functions B_T as the even, odd, and DC basis elements.

The idea behind our *local generative model* is to make explicit the dependence of the quality of approximation in (23) on the distance x from the central point $x_0 = 0$. Low-dimensional parametric models can only locally model nonstationary signals; therefore, observations far from x_0 should neither be attributed to the model nor influence the estimates of the model parameters. This is ignored in related work, e.g., [16], [33], where using the assumption of White Gaussian Noise, the parameters are derived in terms of the Least Squares Error (LSE) projection of the signal on a linear basis.

We account for locality by associating a *confidence* value $G(x)$, e.g., a Gaussian function, with the model-based predictions at point x, as shown in Fig. 6. The quantity $G(x)$ decreases with the distance from the point $x_0 = 0$ and a background model is introduced to model appearance away from x_0. This can be formalized using a binary random variable z_x to indicate whether the observation at x is due to the foreground or background hypothesis, with expectation $P(z_x = 1|x)$ equaling the confidence value $G(x)$ there. The likelihood of the observation $O(x)$ at point x for a set of parameters $\mathbf{E}$ then writes

$$P_T(O(x)|x, \mathbf{E}) = \sum_{z_x=\{0,1\}} P_T(O(x), z_x|x, \mathbf{E}) \quad (27)$$

$$= \sum_{z_x=\{0,1\}} P_T(O(x)|z_x, x, \mathbf{E}) P(z_x|x) \quad (28)$$

$$= \underbrace{P_{T,f}(O(x) - I_T(x|\mathbf{E}))G(x)}_{z_x=1} + \underbrace{P_{T,b}(O(x))(1 - G(x))}_{z_x=0}. \quad (29)$$

The foreground distribution $P_{T,f}$ uses the model synthesis $I_T(x|\mathbf{E})$ to explain the observation $O(x)$, while we use a uniform background distribution on intensity $P_{T,b}(O(x)) = c_B$. Note that $P_T(z_x|x, \mathbf{E}) = P(z_x|x)$ since the confidence in the class prediction depends only on the location x.

A merit of this approach is that it conceptually disentangles the synthesis from the likelihood expressions; this facilitates the construction of simple bases for idealized signals, like sinusoids. Locality is taken into account separately, allowing for increased flexibility compared to previous work [16], [33], where the basis elements have spatially decaying envelopes in order to simultaneously account for signal properties and model locality.

3.2.2 Likelihood Expression and Parameter Estimation

For independent errors, we have

$$\log P_T(O|\mathbf{E}) = \sum_x \log P_T(O(x)|x, \mathbf{E}), \quad (30)$$

which is in general hard to optimize with respect to the model parameters since the summation of (29) appears inside the logarithm. We therefore derive a tractable lower bound of (30) using the concavity of the $\log$ function and Jensen's inequality:

$$LB_T(O) = \sum_x G(x) \log P_{T,f}(O(x)|x) + \sum_x (1 - G(x)) \log c_B. \quad (31)$$

In the following, this lower bound will be used instead of the original expression for the data likelihood. Under the assumption of WGN with variance σ^2, the first term of (31) writes

$$\sum_x G(x) \log P_{T,f}(O(x)|x) = -\frac{1}{2\sigma^2}\mathcal{R} - \sum_x G(x) \log(\sqrt{2\pi}\sigma), \quad (32)$$

$$\mathcal{R} = \sum_x G(x)[O(x) - I_T(x)]^2 \overset{(24)}{=} [\mathbf{O} - \mathbf{BE}]^T \mathbf{G}[\mathbf{O} - \mathbf{BE}], \quad (33)$$

where $\mathbf{G}$ is an $N \times N$ diagonal matrix, with $\mathbf{G}_{i,i} = G(i)$, $\mathbf{O}$ is the $N \times 1$ observation matrix, and $\mathbf{B}$ and $\mathbf{E}$ are $N \times D_T$ and $D_T \times 1$ matrices, respectively, expressing (24) concisely in matrix notation. Apart from the weighted reconstruction error $\mathcal{R}$, the rest of the terms in (32) are constant, so the maximum condition for (32) yields the Weighted LSE solution:

$$\hat{\mathbf{E}} = \mathbf{D}^{-1}(\mathbf{B}^T\mathbf{GO}), \quad \mathbf{D} = \mathbf{B}^T\mathbf{GB}, \quad (34)$$

$$\mathcal{R} = \mathbf{O}^T\mathbf{GO} - \hat{\mathbf{E}}^T\mathbf{D}\hat{\mathbf{E}}. \quad (35)$$

The dependence of the reconstruction error on the model can thus be expressed using only the parameters $\hat{\mathbf{E}}$: The first term of (35) is model independent, while the second involves only the parameters and a fixed matrix.

If the confidence function $G(x)$ is a Gaussian, the product of the three basis elements in (24) with $G(x)$ yields the impulse responses of a Gaussian, an even and an odd-symmetric Gabor filter. This means that the convolution with these three filters furnishes at each pixel the values of the 3×1 vector $\mathbf{B}^T\mathbf{GO}$. Further, considering the DC response of the even Gabor filter to be negligible, we have

$$\sum_x G(x)\cos(x) \simeq 0 \quad \rightarrow \quad \mathbf{D} = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{bmatrix}, \quad (36)$$

$$\sum_x G(x)\cos(2x) \simeq 0 \quad \rightarrow \quad D_1 = D_2 = \frac{1}{2}\sum_x G(x), \quad (37)$$

meaning that the inversion of the matrix in (34) amounts simply to a normalization of the Gabor/Gaussian filter responses. Introducing

$$G'(x) = \frac{G(x)}{\sum_x G(x)} = \frac{1}{\sqrt{2\pi}\sigma_{GB}} \exp(-\frac{x^2}{2\sigma_{GB}^2}), \quad (38)$$

where σ_{GB} is the spread of the Gabor filter, we have the following ML estimates:

$$\hat{E}_1 = 2\sum_x G'(x)\sin(x)O(x), \quad (39)$$

$$\hat{E}_2 = 2\sum_x G'(x)\cos(x)O(x), \quad (40)$$

$$\hat{E}_3 = \sum_x G'(x)O(x). \quad (41)$$

These expressions constitute the first main result of this section: filtering with an even/odd Gabor filter is interpreted as estimating the optimal weighted projection of the signal on the corresponding sinusoidal basis element. The weighting guarantees that points closer to the center of the filter are reconstructed more accurately than the ones further away; in the same sense, filtering with a Gaussian function performs a weighted projection on a constant basis. This also justifies the use of normalized convolution [26], [54] to a broad set of problems, including Gabor analysis here and the use of Quadrature Filter Pairs (QFPs) in Section 3.3.

Further, the decrease in reconstruction error due to the model of (23) can be expressed as

$$\hat{\mathbf{E}}^T\mathbf{D}\hat{\mathbf{E}} = \sum_x G(x)\left[\frac{\hat{E}_1^2+\hat{E}_2^2}{2}+\hat{E}_3^2\right] = \sum_x G(x)\left[\frac{\hat{A}^2}{2}+\hat{B}^2\right].$$

Since $\hat{E}_3 = \hat{B}$ is determined by the Gaussian filter, the only term depending on the Gabor filter's orientation is $\hat{E}_1^2 + \hat{E}_2^2 = \hat{A}^2$, which, by (19), amounts to the square of the amplitude-based channel selection criterion. Using two constants c_1 and c_2 that are independent of the filter's orientation, the lower bound (31) can be written as

$$LB_T(O) = c_1\hat{A}^2 + c_2. \quad (42)$$

This constitutes the second main result of this section: We interpret amplitude-based channel selection as multiple hypothesis testing since we choose the channel that maximizes the lower bound on the observations' log likelihood given the underlying model assumption of (23). Extending this result, we provide in Appendix B a probabilistic justification for the Teager Energy channel selection criterion under the assumption of Brownian noise.

3.3 Edge and Texture Discrimination

Edge detection is now phrased in the same generative model setting, thereby making it possible to evaluate the edge and texture hypotheses on common grounds. Specifically, the decrease in the weighted reconstruction error induced by each of the two hypotheses serves as a means to decide which type of structure is dominant. We can thus estimate their posterior probabilities, which, as we subsequently show, can be used for cue combination in image segmentation.

3.3.1 Edge Synthesis Equations

In [39], it was argued that edges are perceived at locations of *phase congruency*, where the signal is approximated by a Fourier series expansion of the form:

$$O(x) \simeq I_E(x) = A\sum_k a_k\cos(\omega_0 kx + \phi_o) + B. \quad (43)$$

Above, I_E is the approximation to the signal using the edge class model and ω_0 is the fundamental frequency of the Fourier series. The phase offset ϕ_o is common to all harmonic components; hence, at point $x_0 = 0$, phase congruency occurs with different values of ϕ_o, yielding different types of edges, e.g., for $\phi_o = \pi/2$ and $\phi_o = 0$, step-like and bar-like edges are perceived, respectively.

The use of QFPs for edge detection [40], [45] has been largely motivated by this model: QFPs consist of a pair of even and odd-symmetric filters h_e and h_o, with zero DC response and identical Fourier spectrum magnitude. Convolution of a signal of the form of (43) with a QFP and estimation of the *local energy* [40] measurement $LE(x) = (h_e * O)^2 + (h_o * O)^2$ yields a quantity invariant to ϕ_o, while, in [45], it is proven that such an approach can better localize composite edges than Canny edge detection.

We can bring (43) in the previous setting by writing it as an expansion on a linear basis:

$$I_E(x|\mathbf{E}) = \sum_{i=1}^{D_E} E_i B_{E,i}(x), \quad \text{where} \quad D_E = 3, \quad (44)$$

$$E_1 = A\cos(\phi_o), \quad E_2 = -A\sin(\phi_o), \quad E_3 = B, \quad (45)$$

$$B_{E,1}(x) = \sum_k a_k\cos(\omega_0 kx), \quad (46)$$

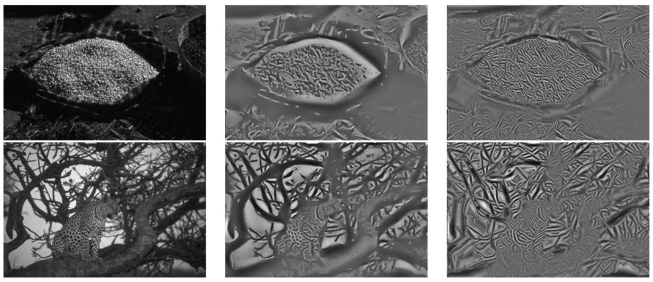
$$B_{E,2}(x) = \sum_k a_k\sin(\omega_0 kx), \quad (47)$$

and $B_{E,3}(x) = 1$. Obviously, the basis elements are periodic, which is in contrast with the notion of an edge. However, as shown in Fig. 6b and detailed in Appendix A, the spatial support of the confidence function is so small that it takes into account less than half of their period, where the basis elements behave like typical edge signals. Consequently, the filters $h_e = G(x)B_{E,1}(x)$ and $h_o = G(x)B_{E,2}(x)$ are tuned to edge signals.

Considering negligible the spectrum magnitude of $h_e + jh_o$ for negative frequencies, we can view h_e and h_o as a Hilbert pair. Following the same steps as in Section 3.2, we can then show that filtering with this pair amounts to estimating the optimal weighted projection of the image neighborhood onto the edge-function basis, $B_{E,1}$, $B_{E,2}$. Further, its squared local energy $(h_e * I)^2 + (h_o * I)^2$ locally measures the fitness of the edge model to the observed data.

Comparing the generative models for edges and textures, the model of (23) accounts for signals that are ideally localized in frequency, while that of (43) can account for sharp transitions, using higher frequency terms. Further, in the texture-model case, as shown in Fig. 6a, the support of the related confidence functions allows for several oscillations of the harmonic component, thereby rendering their outputs better tuned to periodic signals than isolated intensity variations.

A picture of what each class "sees" in an image can be obtained by reconstructing the image in terms of the locally estimated model parameters; in Fig. 7, we see that the edge



(a) (b) (c)

Fig. 7. Reconstruction of the nonsmooth component of the image in (a) using (b) the edge model of (43) and (c) the texture model of (23). The reconstruction in image (c) favors the periodic aspects of the local structure, contrary to (b), where sharp nonperiodic transitions in intensity are more pronounced.

model of (43) interprets image variation in terms of sharp transitions in intensity, while the texture model of (23) focuses on oscillatory image patterns. The contrast of the reconstructed images is proportional to the amplitude of the model or, equivalently, to the log likelihood of the image patches given the models; one can observe that, along object borders, the edge contrast is sharper than that of the texture model and vice versa in the interior of textured areas. More extensive results are available at the first author's web page.

3.3.2 Null Hypothesis—Smooth Signals

Complementing the set of models, smooth (S) regions are modeled as constant signals:

$$O(x) \simeq I_S(x) = B. \tag{48}$$

This model reduces the credibility of texture/edge features at smooth image areas; using the same rationale as with the previous two models, we can view the output of convolution with a Gaussian function as the optimal weighted projection of the observed image data on the basis element $\mathbf{1}$. This model is bound to result in a larger reconstruction error than the other two since it arises as a special case of (23) and (43) by setting $A = 0$. We therefore introduce a Minimum Description Length (MDL) penalty term equal to $\mathrm{MDL} = -\frac{n}{2}\log(\sum_x G(x))$; we set $n = 2$ equal to the additional number of parameters, while the term $\frac{1}{2}\log(\sum_x G(x))$ can be derived by adapting to our case the proof in [24, volume 2, chapter 6, appendix F].

3.3.3 Edge versus Texture Discrimination over Multiple Scales

Up to now, it has been assumed that all three models attempt to explain an image neighborhood at a fixed scale; since we choose among the considered classes by combining evidence from all scales, we need to render comparable likelihood terms that account for a different number of image observations.

First, we derive a quantity invariant to additive constants by considering the term $\mathcal{G}_\mathcal{H} = \log\frac{P(O|\mathcal{H})}{P(O|S)}$, with $\mathcal{H}$ being one of E and T and S being the smooth hypothesis. Replacing the log likelihoods with their lower bounds then yields

$$\mathcal{G}_\mathcal{H} = \frac{1}{2\sigma^2}\left(\mathbf{E}_\mathcal{H} D_\mathcal{H} \mathbf{E}_\mathcal{H}^T - \mathbf{E}_S D_S \mathbf{E}_S^T\right) + \mathrm{MDL}(\mathcal{H}). \tag{49}$$

The first summand quantifies the increase in log likelihood gained by introducing the more complex model $\mathcal{H}$, while $\mathrm{MDL}(\mathcal{H}) = -\log(\sum_x G_\mathcal{H}(x))$ favors the smooth hypothesis at areas where this gain is not significant; the subscript $\mathcal{H}$ has been added here to G since different hypotheses may choose different scales and, thus, different confidence functions G. Apart from $\mathrm{MDL}(\mathcal{H})$, the quantity $\mathcal{G}_\mathcal{H}$ scales proportionally to the area under the confidence function $G_\mathcal{H}$. Therefore, by dividing it with $c\sum_x G_\mathcal{H}(x)$, we obtain an approximately scale-invariant quantity, namely, the per-pixel gain in log likelihood:

$$\mathcal{E}_\mathcal{H} = \frac{\mathcal{G}_\mathcal{H}}{c\sum_x G_\mathcal{H}(x)}, \tag{50}$$

which is used subsequently instead of $\mathcal{G}_\mathcal{H}$ to decide whether the observation is smooth or is better modeled by $\mathcal{H}$. The deviation from scale invariance is due to the MDL-related term $-\frac{\log(\sum_x G_\mathcal{H}(x))}{\sum_x G_\mathcal{H}(x)}$ and shows a small preference to hypotheses that model larger image areas, which is intuitively plausible. Two design parameters in this final expression are the factor c that determines the crispness of the subsequent decisions and the assumed variance of the noise process σ. Eventually, these parameters should be estimated using ground-truth data, but we have obtained satisfactory results on a large variety of images using the constant values $\sigma = 10^{-1}$ and $c = 10$ for images taking values in [0, 1].

Having derived this criterion allows us to compare the two different classes, even if they choose submodels residing at different scales: Separately finding the maxima of the quantities $\mathcal{E}_E$ and $\mathcal{E}_T$ over scales and orientations, we can write, e.g., for the texture-class posterior:

$$P(T|O) = \frac{P(O|T)}{P(O|T) + P(O|E) + P(O|S)} = \frac{R_T}{R_T + R_E + 1},$$
$$\text{where} \qquad R_T = \frac{P(O|T)}{P(O|S)} = \frac{1}{1 + \exp(-\mathcal{E}_T)}, \tag{51}$$

and, similarly, $R_E = 1/(1 + \exp(-\mathcal{E}_E))$. The posterior for the edge class is derived in an analogous manner, while, for the smooth class, it is derived from the residual: $P(S|O) = 1 - P(E|O) - P(T|O)$.

This approach has been applied on a variety of natural images, like those in Figs. 8 and 9, giving plausible results.[1] In these images, the probability of an edge is typically higher along the borders of objects, while, at textured regions, it is lower than what would be indicated by a direct application of an edge detection filter. We also observe that, on object borders, there is a decrease in the probability of texture since the edge model explains away the intensity variation.

We thus see that the generative model approach allows us to interpret probabilistically and attach likelihood and posterior probability terms to filtering operations; as we show next, these can be practically incorporated in subsequent tasks like image segmentation. Finally, based on the generative model setting, we can also address

1. We distribute our Matlab code at http://cvsp.cs.ntua.gr/software.

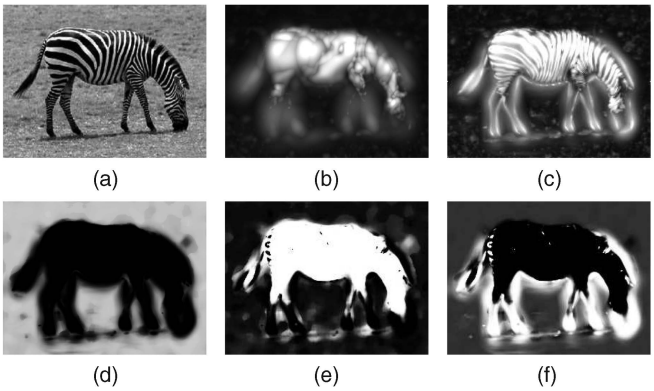


Fig. 8. Discrimination of textured areas from edges, using local generative models: (a) input image, (b) texture, and (c) edge model amplitude, $\sqrt{A_1^2 + A_2^2}$, with A_1 and A_2 estimated according to (34). (d)-(f) Posterior probabilities for the three classes considered, using (51). (a) Input image. (b) Texture amplitude. (c) Edge amplitude. (d) Prob(smooth). (e) Prob(texture). (f) Prob(edge).

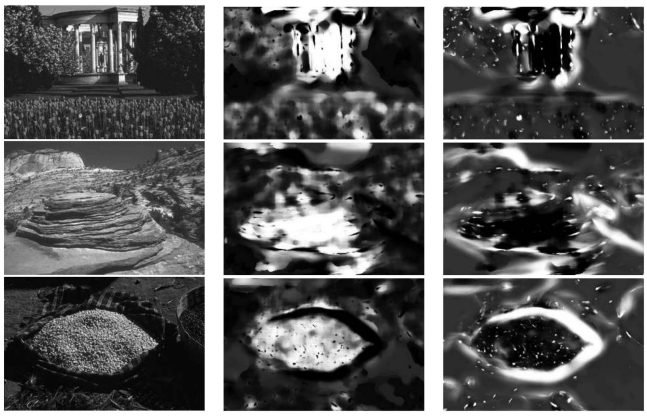


Fig. 9. (a) Input images and posterior probabilities of (b) texture and (c) edge models, respectively.

problems like boundary effects, missing data, and the nonzero DC response of even-symmetric filters. For the first two problems, we assign zero weight to the reconstruction error at locations with missing data and points beyond the image border, thereby getting responses with minimal boundary effects. Concerning the last problem, $D = [\mathbf{B}^T\mathbf{G}\mathbf{B}]$ is not diagonal and its inverse is computed to derive the optimal even/odd projection coefficients. All of the matrix inversions are computed offline, resulting in minimal additional computational burden. Generally, the pair of even/odd symmetric filters does not need to constitute a Hilbert pair, as, e.g., in [40], since the expression of the weighted reconstruction error does not in any way constrain the basis elements used.

4 Unsupervised Texture Segmentation with AM-FM Features and Weighted Curve Evolution

In this section, after a short review of existing region-based curve evolution methods for texture segmentation, we present our contributions comprising 1) the use of AM-FM features as cues for variational unsupervised texture segmentation and 2) the introduction of Weighted Curve Evolution (WCE) as a method that allows us to combine texture, edge, and intensity cues in a locally adaptive manner. Experimental results using images from the Berkeley data set [38] are provided at the end of the section, validating the merits of our contributions on the segmentation task.

4.1 Curve Evolution Methods for Texture Segmentation

The first variational region-based textured image segmentation techniques [30], [57] used modified versions of the Mumford-Shah functional [41] to incorporate the multidimensional features used for texture description. Most recent algorithms [5], [7], [44], [49], [55] use similar energy criteria and rely on curve evolution for energy minimization.

A significant precursor of current work has been the Region Competition method [57], which introduced probabilistic information in curve evolution while clarifying and unifying different variational criteria. The idea of this algorithm is the maximization of the probability of the features F extracted from an image, using a set of M regions R_i, within which the observations are assumed to follow a simple region-specific parametric distribution with parameters θ_i. Combining the log-likelihood of the image features with a prior term on the length of region borders $\mathbf{\Gamma} = \{\Gamma_1, \ldots, \Gamma_M\}$, where $\Gamma_i = \partial R_i$, gives rise to the following functional:

$$J(\mathbf{\Gamma}, \{\theta_i\}) = \sum_{i=1}^{M} \frac{\mu}{2} \int_{\Gamma_i} ds - \iint_{R_i} \log P(F|\theta_i). \qquad (52)$$

Calculus of variations yields the evolution equation for the ith region border:

$$\frac{\partial \Gamma_i}{\partial t} = -\mu\kappa_i \mathcal{N}_i + \log \frac{P(F|\theta_i)}{P(F|\theta_j)} \mathcal{N}_i, \qquad (53)$$

where κ_i is the curvature, $\mathcal{N}_i$ is the outward unit normal of front Γ_i, and j is the neighboring region that competes with i for the observations at the interface position. The front motion steadily increases the functional of (52) by assigning observations to the region that models them better while maintaining the borders smooth at the same time.

In [44], this evolution algorithm has been combined with boundary terms, giving rise to the GAR algorithm:

$$\frac{\partial \Gamma_i}{\partial t} = \left[\lambda \log \frac{P(F|\theta_i)}{P(F|\theta_j)} + (1-\lambda)[-\nabla g(I_E) \cdot \mathcal{N}_i - g(I_E)\kappa_i]\right] \mathcal{N}_i, \qquad (54)$$

where $g(\cdot)$ is a monotonically decreasing function, I_E is the strength of an edge detection system, and λ determines the relative weights assigned to region and edge-based information. The last term is inspired from the Geodesic Active Contour model [6], [25] and forces the region borders to stay close to edges. Further, for the implementation of this evolution law, the use of the level set technique [42] was proposed; this has been established as an elegant mathematical tool for the solution of problems involving evolving interfaces, offering robustness, topological flexibility, and mathematical tractability.

In our implementation, which follows [44], we use a separate embedding function for each front and solve the evolution equation using an explicit integration scheme. For efficiency and robustness, a multiresolution scheme in combination with the narrow-band method is used, while, for the reinitialization of the embedding functions, we use the method in [12]; more efficient schemes could be used, however, as, e.g., in [43].

4.2 Modulation Features for Texture Segmentation

In the supervised texture segmentation scenario, e.g., [44], the high dimensionality of texture features derived from a Gabor filterbank can be dealt with by choosing the channels that maximally separate different textures. It is, however, harder to tackle the unsupervised problem, where heuristics are typically used. In a recent attempt to face the high-dimensionality problem [49], a vector valued diffusion procedure has been used to smooth a compact texture descriptor, derived from local image derivatives:

$$\frac{\partial u_i}{\partial t} = \operatorname{div}\left(g(\sum_{k=1}^{3} |\nabla u_k|)\nabla u_i \right), \quad g(x) = \frac{1}{|\nabla u| + 10^{-3}}, \quad (55)$$

where the channels are initialized to $u_1(\cdot,0) = I_x^2$, $u_2(\cdot,0) = I_x I_y$, $u_3(\cdot,0) = I_y^2$. Combined with image intensity, the resulting four-dimensional feature vector offers satisfactory results for the unsupervised segmentation of textured images. A problem mentioned in [49] is that these features do not carry information about texture scale, but, in [5], a diffusion-based scale measure has been introduced to extend the original feature vector. Still, when using the output of a nonlinear diffusion procedure, one does not have a firm understanding of what the features stand for since they are a "by-product" of a complex process.

We argue that the DCA features are better suited for segmentation for primarily three reasons: First, information about texture scale, contrast, and orientation is captured with a low-dimensional descriptor. Second, the generative model interpretation facilitates the estimation of the posterior probability of the texture and edge classes, which we use subsequently to determine the influence of the texture features on the evolution process. Third, DCA features are easily interpretable: They are derived from a model for texture representation and can synthesize the textured component of the image, thereby showing us what the model "sees" in the image.

The use of DCA features for image segmentation was initially proposed in [48], [51] and subsequently in [11], [28], where results with various textured images were demonstrated for proof of concept. The feature vector consists of the image intensity and the DCA components, namely, amplitude, frequency magnitude, and orientation. The distribution $P(\cdot|\theta_i)$ within region i is modeled as a product of a multivariate Gaussian for the first three features and a von Mises distribution for the orientation feature θ; details on the latter are given in Appendix C. Parameter estimation for the distributions $P(\cdot|\theta_i)$ is performed in alternation with curve evolution, yielding an adaptive unsupervised image segmentation scheme. This contrasts with the initial work in [48], combining curve evolution techniques with DCA features: Therein, curve evolution solely using a Geodesic Active Contour term was used at postprocessing to simplify the borders by introducing curvature information. In this work, arbitrary initial conditions are used for the curves, which automatically detect the image segments.

4.3 Cue Combination via Weighted Curve Evolution

When the assumptions underlying feature extraction do not correspond to the signal behavior, the features are meaningless and can mislead the segmentation. For example, on object borders, texture features are erroneously active; at smooth regions, the orientation features behave erratically; while, in the interior of textured regions, the edge and intensity cues can impede the evolution of the curve. Recent approaches to dealing with this problem include [16], [31], where fairly intricate techniques are used to determine the textured areas in an image. Instead, using a generative model approach, one can quantify which hypothesis is most reliable at a specific location in the image, thereby assigning a confidence measure to each of the modalities used for cue extraction. In a manner inspired from the fusion literature, we incorporate the class posterior probabilities derived in (51) in the Region Competition algorithm, thereby rendering the curve evolution immune to the aforementioned problems.

Supra-Bayesian fusion methods [21] combine the outputs of multiple classifiers ("experts") by treating them as random variables and considering their joint distribution. Specifically, for two classes i and j and a set of features F, any soft classifier using these features provides a *log-odds* quantity $\mathcal{L}_F = \log\frac{P(F|i)}{P(F|j)}$. This quantity is treated as a random variable and is assumed to follow a Gaussian distribution conditioned on the actual class of the features:

$$P(\mathcal{L}_F|i) \sim N(\mu_i, \sigma^2), \quad P(\mathcal{L}_F|j) \sim N(\mu_j, \sigma^2), \quad (56)$$

where σ is common for both cases. These distributions quantify the certainty associated with any classifier decision: A large σ indicates a low confidence in log-odd accuracy and diminishes the effect of a large value of $\mathcal{L}_F$. For a good classifier, if the data F are due to hypothesis i, it is highly probable (low σ) that $\mathcal{L}_F$ will take a high value ($\mu_i \gg \mu_j$) and vice versa.

The results of N classifiers using different features or different classification methods can be integrated based on this approach. Their log-odds $\boldsymbol{\mathcal{L}} = [\mathcal{L}_1, \ldots, \mathcal{L}_K]$ are viewed as a multidimensional random variable that follows a Gaussian distribution conditioned on the class i:

$$P(\boldsymbol{\mathcal{L}}|i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \quad P(\boldsymbol{\mathcal{L}}|j) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}). \quad (57)$$

The posterior log-likelihood ratio given all the expert odds then equals

$$\log\frac{P(i|\boldsymbol{\mathcal{L}})}{P(j|\boldsymbol{\mathcal{L}})} = (\boldsymbol{\mathcal{L}} - \frac{\boldsymbol{\mu}_i + \boldsymbol{\mu}_j}{2})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j). \quad (58)$$

For the special case where the classifiers behave independently, we have a diagonal covariance matrix; further, by appropriately scaling and shifting the classifier outputs, we can guarantee that $\boldsymbol{\mu}_i = -\boldsymbol{\mu}_j$ and all the elements of $\boldsymbol{\mu}_i$ equal unity so that we have

$$\log\frac{P(i|\boldsymbol{\mathcal{L}})}{P(j|\boldsymbol{\mathcal{L}})} = \sum_{k=1}^{K} \frac{\mathcal{L}_k}{\sigma_k^2}. \quad (59)$$

This formula expresses a straightforward idea: When a classifier gives noisy results, i.e., has a large σ, a lower weight should be assigned to its decision and vice versa. The use of log-odds is particularly convenient in that we express the combination operation in terms of a summation.

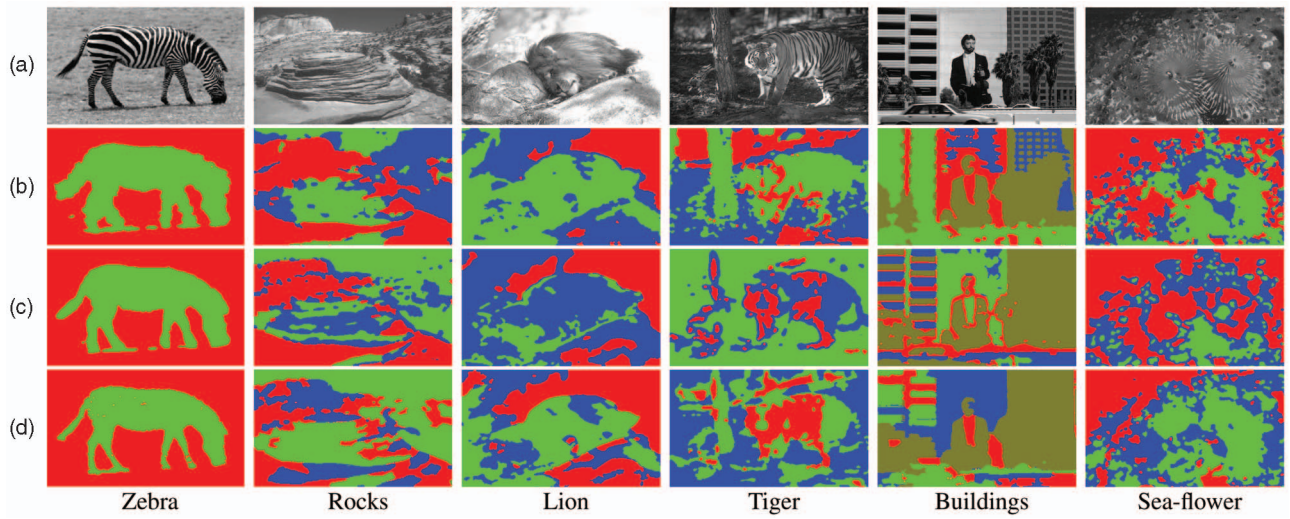


Fig. 10. Segmentation results using different features and evolution schemes. (a) Input image. (b) Segmentation using DCA features and intensity, as in our initial work [28]. (c) Segmentation using our implementation of the diffusion features in [49]. (d) Segmentation using DCA features in conjunction with the Weighted Curve Evolution method.

Within this framework, the probabilistic balloon force of (53),

$$\mathcal{L}_F = \log \frac{P(F|\theta_i)}{P(F|\theta_j)},$$

can be seen as classifying the observed features F into either region i or region j and advancing the fronts so as to correctly classify the observations. The link we build consists of treating the class posterior probabilities of (51) as indicative of the accuracy of a classifier that uses a subset of these features. We consider that three classifiers are used to determine the evolution of the curve: one based on intensity, another on texture cues, and a loosely defined "context-based" classifier that groups pixels into separate classes across edges; the term $(-\nabla g \cdot \mathcal{N})\mathcal{N} - g\kappa\mathcal{N}$ in (54) has this effect close to edges, i.e., aligning the segment borders with the edges. For the first two, we have, using the previous notation,

$$\mathcal{L}_T = \log \frac{P(F_T|\theta_{T,i})}{P(F_T|\theta_{T,j})}, \quad \mathcal{L}_S = \log \frac{P(F_S|\theta_{S,i})}{P(F_S|\theta_{S,j})}, \tag{60}$$

where, by $P(F_c|\theta_{c,i})$, we denote the likelihood of the feature set F_c related to class c (texture-T or smooth-S) under the distribution P of region i, whose parameters are $\theta_{c,i}$. For the final decision, each expert's decision is weighted by the posterior probability of each hypothesis (51); this way, for textured areas, the texture features have a larger impact on the evolution of the curve than the intensity features and vice versa for smooth regions. Equation (54) then gives rise to an evolution scheme that we call *Weighted Curve Evolution*:

$$\frac{\partial \Gamma_i}{\partial t} = \left[\sum_{c \in T,S} w_c \log \frac{P(F_c|\theta_{c,i})}{P(F_c|\theta_{c,j})} + w_E[(-\nabla g \cdot \mathcal{N}) - g\kappa] \right] \mathcal{N}. \tag{61}$$

In (61), we use as weights w_E, w_T, and w_S the posterior probabilities of the edge, texture, and smooth hypotheses. Note that weighting with w_E does not interfere with the edge detection procedure used to estimate ∇g since it only indicates the importance of the edge information.

We have deliberately used w_E, w_T, and w_S and avoided the posterior probability notation to clarify that apart from the intuitive motivation presented earlier, no formal connection is implied between the class posterior probabilities and an optimal combination of the results. Any other measure of "texturedness," "edgeness," and "smoothness" could be used instead, like the ones examined in [31], [37], [53], while our measures of these quantities could be equally well used in the context of their work.

4.4 Experimental Evaluation

We have applied our method to the unsupervised segmentation of a large set of natural images, including the Berkeley Segmentation Benchmark [38]. To demonstrate the merit of our contributions, we have considered alternative features and segmentation approaches, as well as a simpler version of our approach, initially presented in [28].

Initially, in Fig. 10, we present segmentation results using three alternative schemes: In Fig 10b, the results of curve evolution along [28] are shown, using the 3D DCA texture descriptor obtained with energy-based channel selection. In Fig 10c, we use the nonlinear diffusion-based feature set in [49], using our own implementation of the diffusion scheme. Finally, in Fig 10d, we show results obtained with the DCA features and the cue integration scheme described in Section 4.3. For all of the results that we show, the number of fronts is manually determined, while similar segmentations are obtained for different numbers of fronts. In general, we consider the task of estimating the number of segments a high-level task, while our concern in this work is improving the performance of segmentation as a front-end module.

Concerning our original approach [28], we observe the problems mentioned earlier: Even though satisfactory results are obtained with heavily textured images as, e.g., in the tiger and sea-flower scenes, on smooth areas the erratically behaving orientation vector leads to wiggly boundaries. Further, for example, for the zebra image, the large texture amplitude estimate at the object's borders inflates the foreground region since the competing background hypothesis is

TABLE 2
Average/Median Values of BCE for Different Segmentation Schemes and Front Numbers (Lower Is Better)

Fronts	Optimal	2	3	4	5	6	7	8	9	10
DCA, WCE	.38/.39	.46/.49	.49/.51	.51/.52	.54/.53	.54/.53	.57/.57	.59/.58	.59/.59	.61/.61
DCA, Plain	.39/.39	.48/.51	.50/.51	.51/.52	.54/.53	.56/.56	.58/.58	.60/.59	.62/.61	.63/.61
[49], WCE	.40/.41	.47/.49	.51/.53	.52/.53	.55/.55	.57/.58	.59/.58	.60/.60	.62/.61	.63/.63
[49], Plain	.40/.42	.48/.50	.52/.53	.55/.54	.56/.56	.59/.59	.60/.60	.63/.62	.63/.63	.65/.65
N. Cuts	.41/.43	.49/.51	.52/.53	.55/.53	.55/.55	.59/.60	.60/.59	.63/.61	.63/.63	.64/.65

characterized by low-amplitude features and, hence, does not easily explain these observations.

The features in [49] perform comparably to the DCA features but suffer from the same problems on smooth areas. In many figures, e.g., the rocks, lion, and tiger images in Fig. 10, the borders of objects are grouped in a separate region since the image derivatives due to object borders are treated as texture features and call for a region to explain them.

The Weighted Curve Evolution results are immune to the confusion between edge and texture cues; better segmentation results are typically obtained, with the region borders accurately localizing the object borders. The negative effects of orientation features at smooth regions are diminished due to the smaller weight assigned to the texture hypothesis there. One can mention some exceptions, for example, the buildings image in Fig. 10: There, the texture cue for the left building is not strong enough and leads to its oversegmentation in the Weighted Curve Evolution case.

Coming to systematic evaluation results on the Berkeley Benchmark, the first segmentation evaluation measure we considered was the Bidirectional Consistency Error (BCE) introduced in [36]. This quantifies, in a smooth manner, the overlap between a machine-generated segmentation S_M and a set of human segmentations $S_{1,\dots K}$, by comparing, at each pixel p_i, the machine-generated segment $R_M(p_i)$ to all possible human-generated segments $R_{1\dots K}(p_i)$ containing p_i. Its expression is

$$BCE(S_M) = \frac{1}{n}\sum_{i=1}^{n} \min_k \max(E(S_M, S_k, p_i), E(S_k, S_M, p_i)), \tag{62}$$

$$\text{where } E(S_M, S_k, p_i) = \frac{|R_M(p_i) \setminus R_k(p_i)|}{|R_M(p_i)|}, \tag{63}$$

which averages, over the n image pixels, a symmetric segmentation discrepancy measure defined in terms of $E(S_M, S_k, p_i)$. The latter measures the degree to which region R_M is not contained in region R_k via normalized area differences. It is shown in [36] that the BCE sharply separates human from random segmentations and we therefore use it here as a region-based segmentation evaluation measure.

Another measure used for segmentation evaluation in [13] is the F-measure related to the Precision-Recall measurements obtained by treating the segmentation boundaries as an edge map. Precision equals the proportion of correctly detected edges and Recall is the proportion of human-generated boundaries that were detected. The F-measure equals their geometric mean and summarizes a detector's performance. Even though this measure ignores the connected components provided by a segmentation, it quantifies the accuracy of the segment boundary locations.

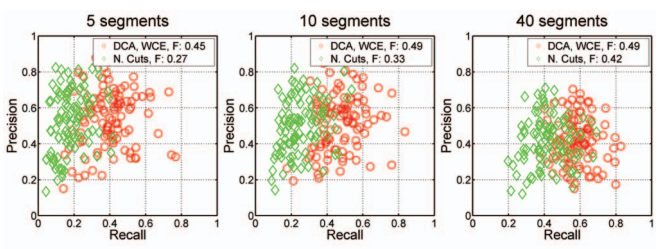
We use both measures since they behave complementarily. The F-measure severely penalizes missed boundaries and can improve for oversegmentations of the image, while BCE favors a moderate number of large segments.

To explore the merits of our approach, we have compared our method to the alternative segmentation methods on the Berkeley Benchmark. The diffusion features are extracted using our own implementation, designed to accurately match the results in [49]. The Normalized Cuts' results have been obtained using the Berkeley Segmentation Engine, using gray-scale images segmented with texture and intensity cues.

We compute the results obtained for an increasing number of image regions, without any postprocessing to determine the number of fronts to reduce the different sources of change in performance. For the sake of completeness, we also compute the statistics of BCE that would be obtained if the number of fronts providing the best BCE score was available by an "oracle"; these are shown in the first column of Table 2.

Specifically, in Table 2, we compare the average and median BCE measures obtained from the different segmentation algorithms. The results validate the usefulness of the WCE scheme when comparing the "WCE" with the "plain" columns. Further, the results show a consistent improvement obtained when using our DCA features instead of those in [49] and also when the WCE method is used in conjunction with the latter. Finally, our method is shown to outperform Normalized Cuts for all choices of the number of segments.

In Fig. 11, we use Precision-Recall measurements to further explore the merit of our method compared to Normalized Cuts. We note that the F-measures obtained here are inferior to those reported in [13] since we do not use color in order to compare the results on an equal footing. Even though the Normalized Cut method minimizes a global criterion and can provide a potentially optimal segmentation,



Aggregate F-measure					
Fronts	5	10	20	30	40
DCA, WCE	.45	.49	.50	.49	.48
N. Cuts	.26	.33	.34	.41	.42

Fig. 11. Precision-Recall results on the Berkeley Benchmark obtained for an increasing number of image segments.

Fig. 12. Segmentation results using DCA features and WCE (preferably seen in color).

from our results, we observe that the performance of curve evolution methods is generally better.

5 Conclusion

Motivated by the mathematical clarity and the practical applicability of modulation models, we have pursued their exploitation in texture analysis and segmentation, working at all problem levels. Apart from improvements in feature extraction in itself, where a rigorous contribution to the demodulation problem and an improved channel selection scheme have been presented, our work has focused on the cross-fertilization between ideas from image analysis and computer vision. Starting from establishing a link between AM-FM and generative models, we have phrased Gabor filtering/DCA channel selection in terms of parameter estimation and hypothesis testing, respectively. Building on this link, we have formulated edge detection on common grounds, facilitating the extraction of model-based posterior probabilities for the edge, texture, and smooth classes. Inspired by the fusion literature, we have introduced the Weighted Curve Evolution method that weighs the contributions of different cues according to their model posterior probability. This has allowed the exploitation of these quantities in natural image segmentation, yielding systematically better results, as shown on the Berkeley Segmentation Benchmark.

In future work, we intend to explore the use of AM-FM/DCA models in related problems, like U+V decomposition and image inpainting, as well as the extraction of salient tokens for object detection problems. We believe that AM-FM/DCA models have not yet been brought to their full potential; by combining the transparency of model-based approaches and the efficiency of feature-based approaches, they can serve as a reliable front end for higher level computer vision tasks.

Appendix A

Filterbank Specification

We use isotropic Gabor filters with impulse response $g(x,y) = \exp(-\frac{x^2+y^2}{2\sigma^2}) \exp(ju_c x + jv_c y)$, where σ determines the spatial support and (u_c, v_c) is the localization in the frequency plane. Fifty geometrically scaled filters are used, covering five scales and 10 orientations in the frequency plane, as shown in Fig. 22. The design is similar to [19]; the smallest frequency is $r_o = 0.5$ cycles/image, with the center frequencies $r_c = \sqrt{u_c^2 + v_c^2}$ decreasing geometrically with a ratio $R \approx 0.56$. The central frequency is related to σ as $\sigma \cdot r_c = 3.7$, resulting in a half-peak bandwidth of $B \approx 0.7$ octaves.

For the edge detection filters, we relate to each texture model of the form (23) a corresponding edge model expressed as in (43). The fundamental frequency in (43) is set empirically to $\omega_0 = 0.4\omega_G$, with ω_G being the frequency of the texture model in (23). The Fourier series coefficients a_k in (43) are set equal to those of a periodic square wave.

Appendix B

Brownian Noise and Teager Energy

In Section 3.2.2, a white Gaussian noise (WGN) assumption led to the interpretation of the channel selection criterion as a lower bound on the observation likelihood.

In the more general case of correlated noise, parameter estimation involves diagonalizing the noise covariance matrix [24] and analysis becomes intractable. For Brownian motion, however, one can work on the derivative of the observations that are then modeled in terms of the derivatives of the basis elements plus WGN. This amounts to reconstructing the derivative of the signal O' on the differentiated basis elements, which, specifically for sinusoids, are $B'_{T,1} = -\omega B_{T,2}$ and $B'_{T,2} = \omega B_{T,1}$. The normalization of the basis elements with respect to a multiplying constant is effected by the combination of (34) and (35), so we can consider that the signal derivative is projected onto basis elements $B'_{T,1} = B_{T,2}$ and $B'_{T,2} = -B_{T,1}$. The amplitude A' of the differentiated signal O' can thus be estimated as

$$A' = \sqrt{(O' * GB_{T,1})^2 + (O' * GB_{T,2})^2} \tag{64}$$

$$= \sqrt{(O * GB'_{T,1})^2 + (O * GB'_{T,2})^2} \quad (65)$$

$$= \omega\sqrt{(O * GB_{T,2})^2 + (O * GB_{T,1})^2} = \omega A. \quad (66)$$

Iterating the previous analysis, the data likelihood can be expressed in terms of $(A')^2 = \omega^2 A^2$, where A is the amplitude estimate for the nondifferentiated signal. The quantity $\omega^2 A^2$ equals the response of the energy operator in (4) for a sinusoidal, probabilistically justifying its use for channel selection in (20) in the case of Brownian noise.

APPENDIX C
ORIENTATION DISTRIBUTION

The von Mises density function $P_{VM}(\theta; \theta_0, \kappa) = \frac{\exp(\kappa\cos(\theta-\theta_0))}{2\pi I_0(k)}$ is analogous to the Gaussian for orientational data. I_0 is the Bessel function of the first kind with pure imaginary argument, while the estimator of θ_0 is given by [15] $\theta_0 = \tan^{-1}(\frac{\sum_k \sin(\theta_k)}{\sum_k \cos(\theta_k)})$. κ can be numerically estimated as the solution of

$$I'_0(\kappa) = \hat{\alpha} I_0(\kappa), \;\; \hat{\alpha} = \frac{1}{K}\sqrt{\left(\sum_{k=1}^{K}\sin(\theta_k)\right)^2 + \left(\sum_{k=1}^{K}\cos(\theta_k)\right)^2},$$

with K being the number of observations within the region. Actually, since the orientation estimates range in $[0, \pi)$ and not in $[0, 2\pi)$, we multiply their value by two before inserting them into the parameter estimation and likelihood expressions.

ACKNOWLEDGMENTS

The authors thank G. Papandreou for drawing their attention to the efficient distance transform computation in [12] and D. Dimitriadis for the discussions on 1D Gabor ESA. They wish to thank the reviewers for their constructive comments that helped improve the quality of the paper. This research was supported by the Greek Ministry of Education under program "HRAKLEITOS," the Greek Secretariat for Research & Technology under program "ΠΕΝΕΔ-2001," and the European Network of Excellence "MUSCLE." I. Kokkinos was with the National Technical University of Athens when this paper was first submitted.

REFERENCES

[1] S. Belongie, C. Carson, H. Greenspan, and J. Malik, "Color- and Texture-Based Image Segmentation Using EM and Its Application to Content-Based Image Retrieval," *Proc. Sixth Int'l Conf. Computer Vision,* 1998.

[2] A.C. Bovik, M. Clark, and W. Geisler, "Multichannel Texture Analysis Using Localized Spatial Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 1, 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. Information Theory,* vol. 38, pp. 691-712, 1992.

[4] A.C. Bovik, P. Maragos, and T.F. Quatieri, "AM-FM Energy Detection and Separation in Noise Using Multiband Energy Operators," *IEEE Trans. Signal Processing,* vol. 41, pp. 3245-3265, 1993.

[5] T. Brox and J. Weickert, "A TV Flow Based Local Scale Measure for Texture Discrimination," *Proc. Eighth European Conf. Computer Vision,* 2004.

[6] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic Active Contours," *Int'l. J. Computer Vision,* vol. 22, no. 1, pp. 61-79, 1997.

[7] T. Chan and L. Vese, "Active Contours without Edges," *IEEE Trans. Image Processing,* vol. 10, no. 2, pp. 266-277, 2001.

[8] G.C. Cross and A.K. Jain, "Markov Random Field Texture Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 5, no. 1, pp. 25-39, Jan. 1983.

[9] J. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters," *J. Optical Soc. of America (A),* vol. 2, no. 7, pp. 160-169, 1985.

[10] D. Dimitriadis and P. Maragos, "Robust Energy Demodulation Based on Continuous Models with Application to Speech Recognition," *Proc. Eighth European Conf. Speech Comm. and Technology,* 2003.

[11] G. Evangelopoulos, I. Kokkinos, and P. Maragos, "Advances in Variational Image Segmentation Using AM-FM Models: Regularized Demodulation and Probabilistic Cue Integration," *Proc. Third Int'l Workshop Variational and Level Set Methods,* pp. 121-136, 2005.

[12] P. Felzenswalb and D. Huttenlocher, "Distance Transforms of Sampled Functions," Technical Report TR2004-1963, Cornell CIS, 2004.

[13] C. Fowlkes and J. Malik, "How Much Does Globalization Help Segmentation?" Technical Report CSD-4-1340, Univ, of California Berkeley, 2004.

[14] D. Gabor, "Theory of Communication," *J. IEE,* vol. 93, no. 3, pp. 429-457, 1946.

[15] E. Gumbel, A. Greenwood, and D. Durand, "The Circular Normal Distribution: Theory and Tables," *J. Am. Statistical Assoc.,* vol. 48, no. 261, pp. 131-152, 1953.

[16] C.-E. Guo, S.-C. Zhu, and Y.N. Wu, "A Mathematical Theory of Primal Sketch and Sketchability," *Proc. Ninth Int'l Conf. Computer Vision,* 2003.

[17] C.-E. Guo, S.-C. Zhu, and Y.N. Wu, "Modeling Visual Patterns by Integrating Descriptive and Generative Methods," *Int'l. J. Computer Vision,* vol. 53, no. 1, pp. 5-29, 2003.

[18] J.P. Havlicek, A.C. Bovik, and D. Chen, "AM-FM Image Modeling and Gabor Analysis," *Visual Communication and Image Processing,* Marcel Dekker, 1999.

[19] J.P. Havlicek, D.S. Harding, and A.C. Bovik, "The Multi-Component AM-FM Image Representation," *IEEE Trans. Image Processing,* vol. 5, no. 6, pp. 1094-1100, 1996.

[20] J.P. Havlicek, D.S. Harding, and A.C. Bovik, "Multidimensional Quasi-Eigenfunction Approximations and Multicomponent AM-FM Models," *IEEE Trans. Image Processing,* vol. 9, no. 2, pp. 227-242, 2000.

[21] R. Jacobs, "Methods for Combining Experts' Probability Assessments," *Neural Computation,* no. 7, pp. 867-888, 1995.

[22] A.K. Jain and F. Farrokhnia, "Unsupervised Texture Segmentation Using Gabor Filters," *Pattern Recognition,* vol. 24, no. 12, pp. 1167-1186, 1991.

[23] B. Julesz, "Textons, the Elements of Texture Perception and Their Interactions," *Nature,* vol. 290, pp. 91-97, 1981.

[24] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation and Detection Theory,* vols. 1-2. Prentice Hall, 1993.

[25] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Gradient Flows and Geometric Active Contour Models," *Proc. Fifth Int'l Conf. Computer Vision,* 1995.

[26] H. Knutsson and C.F. Westin, "Normalized Convolution and Differential Convolution," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 1993.

[27] I. Kokkinos, G. Evangelopoulos, and P. Maragos, "Advances in Texture Analysis: Energy Dominant Components and Multiple Hypothesis Testing," *Proc. Int'l Conf. Image Processing,* 2004.

[28] I. Kokkinos, G. Evangelopoulos, and P. Maragos, "Modulation-Feature Based Textured Image Segmentation Using Curve Evolution," *Proc. Int'l Conf. Image Processing,* 2004.

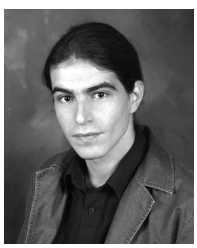
[29] I. Kokkinos and P. Maragos, "A Detection-Theoretic Approach to Texture and Edge Discrimination," *Proc. Fourth Int'l Workshop Texture Analysis and Synthesis,* 2005.

[30] T.S. Lee, D. Mumford, and A. Yuille, "Texture Segmentation by Minimizing Vector Valued Energy Functionals," *Proc. Second European Conf. Computer Vision,* 1992.

[31] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and Texture Analysis for Image Segmentation," *Int'l. J. Computer Vision,* vol. 43, no. 1, pp. 7-27, 2001.

[32] J. Malik and P. Perona, "Preattentive Texture Discrimination with Early Vision Mechanisms," *J. Optical Soc. of America (A),* vol. 7, no. 5, pp. 923-932, 1990.
[33] D. Mallat and Z. Zhang, "Matching Pursuit in a Time-Frequency Dictionary," *IEEE Trans. Signal Processing,* vol. 41, pp. 3397-3415, 1993.
[34] P. Maragos and A.C. Bovik, "Image Demodulation Using Multidimensional Energy Separation," *J. Optical Soc. of Am. (A),* vol. 12, no. 9, pp. 1867-1876, 1995.
[35] P. Maragos, J.F. Kaiser, and T.F. Quatieri, "Energy Separation in Signal Modulations with Application to Speech Analysis," *IEEE Trans. Signal Processing,* vol. 41, no. 10, pp. 3024-3051, Oct. 1993.
[36] D. Martin, "An Empirical Approach to Grouping and Segmentation," PhD dissertation, Univ. of California Berkeley, 2002.
[37] D. Martin, C. Fowlkes, and J. Malik, "Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 5, pp. 530-549, May 2004.
[38] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics," *Proc. Eighth Int'l Conf. Computer Vision,* 2001.
[39] C. Morrone and D. Burr, "Feature Detection in Human Vision: A Phase-Dependent Energy Model," *Proc. Royal Soc. of London B,* vol. 235, pp. 221-245, 1988.
[40] M. Morrone and R. Owens, "Feature Detection from Local Energy," *Pattern Recognition Letters,* vol. 6, pp. 303-313, 1987.
[41] D. Mumford and J. Shah, "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems," *Comm. Pure and Applied Math.,* vol. 42, no. 5, pp. 577-685, 1989.
[42] S. Osher and J. Sethian, "Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations," *J. Computational Physics,* vol. 79, pp. 12-49, 1988.
[43] G. Papandreou and P. Maragos, "Multigrid Geometric Active Contour Models," *IEEE Trans. Image Processing,* vol. 16, no. 1, pp. 229-240, 2007.
[44] N. Paragios and R. Deriche, "Geodesic Active Regions and Level Set Methods for Supervised Texture Segmentation," *Int'l. J. Computer Vision,* vol. 46, no. 3, pp. 223-247, 2002.
[45] P. Perona and J. Malik, "Detecting and Localizing Edges Composed of Steps, Peaks and Roofs," *Proc. Third Int'l Conf. Computer Vision,* 1990.
[46] T.A. Poggio, H. Voorhees, and A. Yuille, "A Regularized Solution to Edge Detection," *J. Complexity,* vol. 4, no. 2, pp. 106-128, Oct. 1988.
[47] J. Portilla and E. Simoncelli, "A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients," *Int'l J. Computer Vision,* vol. 40, no. 1, pp. 49-71, 2000.
[48] N. Ray, J. Havlicek, S. Acton, and M. Pattichis, "Active Contour Segmentation Guided by AM-FM DCA," *Proc. Int'l Conf. Image Processing,* 2001.
[49] M. Rousson, T. Brox, and R. Deriche, "Active Unsupervised Texture Segmentation on a Diffusion Based Space," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2003.
[50] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 888-905, Aug. 2001.
[51] T. Tangsukson and J.P. Havlicek, "AM-FM Image Segmentation," *Proc. Int'l Conf. Image Processing,* 2000.
[52] Z.W. Tu and S.C. Zhu, "Image Segmentation by Data-Driven MCMC," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 657-673, May 2002.
[53] L. Vese and S.J. Osher, "Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing," *J. Scientific Computing,* vol. 19, nos. 1-3, pp. 553-572, 2003.
[54] C.F. Westin, K. Nordberg, and H. Knutsson, "On the Equivalence of Normalized Convolution and Normalized Differential Convolution," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 1994.
[55] A. Yezzi, A. Tsai, and A. Willsky, "A Statistical Approach to Snakes for Bimodal and Trimodal Imagery," *Proc. Seventh Int'l Conf. Computer Vision,* 1999.
[56] S.C. Zhu, Y. Wu, and D. Mumford, "Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling," *Int'l. J. Computer Vision,* vol. 27, no. 2, pp. 107-126, 1998.
[57] S.C. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 9, pp. 884-900, Sept. 1996.

Iasonas Kokkinos received the Diploma in electrical and computer engineering and the PhD degree from the National Technical University of Athens, Greece, in 2001 and 2006, respectively. During 2004, he visited with the Odyssee Team at Sophia-Antipolis, France, and, between 2006 and 2008, he was with the Center for Image and Vision Sciences at the University of California, Los Angeles, as a postdoctoral researcher. He is currently an assistant professor at the École Centrale Paris. His research interests are in the broader fields of signal processing, computer vision, and pattern recognition, and in particular in the application of statistical approaches to the solution of computer vision tasks. He has worked on texture segmentation, biologically motivated vision, and nonlinear dynamical modeling, while his current research focus is on the combination of top-down approaches with bottom-up information for the problems related to object detection. He has served as a reviewer for the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and the *IEEE Transactions on Image Processing* and on the program committees of EMMCVPR '07 and ICCV '07. He is a member of the IEEE.



Georgios Evangelopoulos received the Diploma in electrical and computer engineering and the PhD degree, working with the Computer Vision, Speech Communication, and Signal Processing Group, from the National Technical University of Athens (NTUA), Greece, in 2001 and 2007, respectively. His thesis involved modeling image texture for image decomposition and low-level vision processing. His research interests are in the areas of nonlinear signal processing, visual texture analysis and audio/visual saliency, and generalized event detection, with applications in the fields of image analysis, computer vision, and multimedia processing. He has participated as a researcher for ICCS-NTUA in European research programs and served as a reviewer for the IEEE and OSA. He is a member of the IEEE.

Petros Maragos received the Diploma in electrical engineering from the National Technical University of Athens (NTUA) in 1980, and the MScEE and PhD degrees from the Georgia Institute of Technology (Georgia Tech), Atlanta, in 1982 and 1985, respectively. In 1985, he joined the faculty of the Division of Applied Sciences at Harvard University, Massachusetts, where he worked for eight years as a professor of electrical engineering. In 1993, he joined the faculty of the School of Electrical and Computer Engineering (ECE) at Georgia Tech. During parts of 1996-1998, he was on sabbatical and academic leave, working as the director of research at the Institute for Language and Speech Processing, Athens. Since 1998, he has been working as a professor in the School of ECE at NTUA. His research and teaching interests include signal processing, systems theory, communications, pattern recognition, and their applications to image processing and computer vision, speech processing and recognition, and multimedia. He has served as an associate editor for the *IEEE Transactions on Acoustics, Speech, and Signal Processing* and the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and as an editorial board member for the journals *Signal Processing* and *Visual Communications and Image Representation*. He has served as the general chairman or a cochair of conferences and a member of IEEE DSP committees. His research has received several awards, including the 2007 EURASIP Technical Achievements Award for contributions to nonlinear signal processing and systems theory, image processing, and speech processing. He is a fellow of the IEEE.