

Advances in Variational Image Segmentation using AM-FM models: Regularized Demodulation and Probabilistic Cue Integration.

Georgios Evangelopoulos, Iasonas Kokkinos, and Petros Maragos ^{*}

Computer Vision, Speech Communication and Signal Processing Group
National Technical University of Athens, Greece
{gevag, jkokkin, maragos}@cs.ntua.gr
URL: <http://cvsp.cs.ntua.gr>

Abstract. Current state-of-the-art methods in variational image segmentation using level set methods are able to robustly segment complex textured images in an unsupervised manner. In recent work, [18, 19] we have explored the potential of AM-FM features for driving the unsupervised segmentation of a wide variety of textured images. Our first contribution in this work is at the feature extraction level, where we introduce a regularized approach to the demodulation of the AM-FM -modelled signals. By replacing the cascade of multiband filtering and subsequent differentiation with analytically derived equivalent filtering operations, increased noise-robustness can be achieved, while discretization problems in the implementation of the demodulation algorithm are alleviated. Our second contribution is based on a generative model we have recently proposed [18, 20] that offers a measure related to the local prominence of a specific class of features, like edges and textures. The introduction of these measures as weighting terms in the evolution equations facilitates the fusion of different cues in a simple and efficient manner. Our systematic evaluation on the Berkeley segmentation benchmark demonstrates that this fusion method offers improved results when compared to our previous work as well as current state-of-the-art methods.

1 Introduction

The segmentation of textured images is a long standing problem in computer vision that has been addressed in the framework of variational methods using both boundary- and region-based techniques. The latter are commonly held as more appropriate for this specific problem due to the increased robustness offered by region-based criteria and the difficulty of texture boundary localization.

In the region-based scenario, informative features are used to drive the evolution process; image intensity on its own is a poor cue, since textured images are inherently of varying intensity. Multiband image filtering with filterbanks is commonly used as a preprocessing step that facilitates the extraction of texture information residing at different frequency channels. Even though the outputs of such a filterbank may accurately describe the texture signal, their high dimensionality can lead to suboptimal segmentations.

In recent work [18, 19, 37] the potential of modulation features derived using the AM-FM model of Bovik and coworkers [2, 11, 10, 12] in driving the unsupervised segmentation of textured images has been explored. The low-dimensional texture representation resulting from Dominant Component Analysis (DCA) [10, 12] offers information concerning the local contrast, scale and orientation of the texture signal and can be interpreted as describing the sinusoidal signal that best models a texture locally [18, 20].

^{*} This work was supported in part by the Greek research program HRAKLEITOS, which is co-funded by the European Social Fund (75%) and National Resources (25%), the Greek GSRT research program PENED-2001 and the European NoE MUSCLE

A problem faced by our algorithm is that at smooth areas texture features like orientation are meaningless and introduce erroneous information in the feature vector, while at object borders texture features indicate the presence of a textured region, leading again to suboptimal solutions. Further, the demodulation algorithm used for feature extraction includes high-order image derivatives, that introduce increased noise sensitivity and are not uniquely defined for discrete-time signals.

In this work, our contribution is twofold: first we introduce a regularized version of the algorithm used for feature extraction, involving generalized Gabor filtering and treating errors from inefficient discrete differentiations. Second, we propose a modification of the original Region Competition/ Geodesic Active Regions evolution rule that takes into account the locally estimated confidence in any of the low-dimensional modulation-based features. For this we build upon recent work [18, 20] and provide probabilistic terms quantifying the confidence assigned to the extracted features, relying on a detection theoretic interpretation of the DCA algorithm.

Section II describes previous work and provides the background for later sections. In Section III the regularized demodulation algorithm is described, while Section IV presents the cue integration algorithm proposed. In Section V we demonstrate the merits of using the fused scheme and compare both visually and quantitatively our method's results to those obtained using current state-of-the-art features [38].

2 Previous work: AM-FM Models and Unsupervised Segmentation Methods

2.1 AM-FM Texture Modeling

According to the multicomponent AM-FM model [10], a textured image can be modelled as the superposition of sinusoidal components:

$$I(x, y) = \sum_{k=1}^K a_k(x, y) \cos(\phi_k(x, y)), \quad \omega_k(x, y) = \nabla \phi_k(x, y), \quad (1)$$

where each of the K components is a non-stationary 2-D AM-FM signal, with *instantaneous amplitude* $a_k(x, y)$ and *instantaneous frequency* ω_k . The decomposition of an image in terms of this expression is an ill-posed problem, since one can devise an infinity of AM and FM signals yielding the same image. A separation into well-behaved individual AM-FM components can be accomplished by filtering with a multiband Gabor filterbank [1, 7]; the output of each filter can then be represented as a mono-component AM-FM signal with *narrowband* modulation components, that lend themselves to efficient demodulation algorithms and are intuitively interpretable. Specifically if $f(x, y)$ is a mono-component 2-D AM-FM signal

$$f(x, y) = a(x, y) \cos(\phi(x, y)), \quad (2)$$

its spatially-varying amplitude $a(x, y)$ can be interpreted as modeling local image contrast while the instantaneous frequency vector $\omega(x, y) = \nabla \phi(x, y)$ describes locally emergent spatial frequencies [2, 12]. Efficient estimation of the modulation components of the 2D AM-FM signals can be accomplished via the multidimensional energy separation algorithm [24] that uses a multidimensional energy operator [3]: Let $f(x, y)$ be a twice-differentiable continuous-space real-valued input function. The 2D energy operator Φ is defined by

$$\Phi(f)(x, y) \triangleq \|\nabla f(x, y)\|^2 - f(x, y) \nabla^2 f(x, y) \quad (3)$$

Let now f be a 2D spatial AM-FM signal as in (2). Under certain assumptions on the amplitude and frequency variations [24], applying Φ to f yields the energy product of the squared instantaneous amplitude and frequency magnitude $\Phi[a \cos(\phi)] \approx a^2 \|\omega\|^2$ with an approximation error bounded within a negligible range for locally narrowband signals. Applying Φ to the partial derivatives $f_x = \partial f / \partial x$, $f_y = \partial f / \partial y$ yields the 2D continuous *Energy Separation Algorithm (ESA)* [24]:

$$\sqrt{\frac{\Phi(f_x)}{\Phi(f)}} \approx |\omega_1(x, y)| \quad \sqrt{\frac{\Phi(f_y)}{\Phi(f)}} \approx |\omega_2(x, y)| \quad \frac{\Phi(f)}{\sqrt{\Phi(f_x) + \Phi(f_y)}} \approx |a(x, y)| \quad (4)$$

This algorithm can estimate at each location (x, y) the amplitude and the magnitude of the instantaneous vertical and horizontal frequencies of the spatially-varying 2-D AM-FM signal. The signs of the frequency signals are obtained from the signs of the carriers, approximated by the bandpass filter central frequencies. By replacing the partial derivatives with differences a variety of discrete energy operators emerge. A simple 2D case is:

$$\Phi_d(f)(i, j) = 2f^2(i, j) - f(i-1, j)f(i+1, j) - f(i, j-1)f(i, j+1) \quad (5)$$

Applying Φ_d to a 2D discrete AM-FM signal $f[i, j] = a[i, j] \cos(\phi[i, j])$ yields [24] a nonlinear energy product $\Phi_d[a[i, j] \cos(\phi[i, j])] \approx a^2[i, j](\sin^2(\Omega_1[i, j]) + \sin^2(\Omega_2[i, j]))$, where Ω_1, Ω_2 are the discrete-space instantaneous frequencies. The discrete ESA [24], can give estimates of instantaneous amplitude and frequencies of narrowband image components with an excellent spatial resolution and very low complexity.

2.2 Low-Dimensional AM-FM features via Dominant Component Analysis

The previously described demodulation scheme yields a $3K$ -dimensional feature vector at each point, where K is the number of filters used in the Gabor filterbank. Even though this descriptor offers rich information about the texture signal, it cannot be used as is for segmentation purposes, due to its high dimensionality.

A more compact descriptor can be extracted using the Dominant Component Analysis (DCA) [10, 12] scheme: DCA picks at each image pixel the most active filterbank channel, demodulates its output and uses the resulting AM-FM features to represent the local texture structure. This offers at each image point a three-dimensional feature vector that retains essential information about the texture structure, describing its most prominent characteristics in terms of a sinusoidal signal.

At the heart of the DCA method lies the channel selection criterion used to pick the most active channel at each point; in the original work on DCA the local estimates of the amplitude envelopes A_k for each channel k were used, which are estimated as $A_k = \sqrt{\Im(g_k * I)^2 + \Re(g_k * I)^2}$, where I is the image and g_k is the impulse response of the k -th complex 2-D Gabor filter. In [18] we presented a detection theoretic interpretation of this ‘maximum-amplitude’ channel selection criterion, relating the A_k term with the log-likelihood of the image observations around the neighborhood of each point. Further, it was observed that using an energy-operator-based instead of an amplitude-based selection criterion offers a viable alternative, characterized by better localization accuracy and in [20] this alternative channel selection criterion was cast in a detection-theoretic framework as well.

2.3 Unsupervised Variational Textured Image Segmentation

Region-based techniques are commonly considered as more appropriate for textured image segmentation, since the application of boundary-based techniques and the related variational schemes of Snakes, Deformable models [15, 6] and Geodesic Active Contours [5, 16] is based on the detection of strong variation in texture features [22, 31], which is a non-trivial problem.

Some of the first variational region-based textured image segmentation techniques [21, 46] have used modified versions of the Mumford-Shah functional [28] appropriately modified to incorporate the multi-dimensional features used for texture description; building upon this work, current state-of-the-art algorithms in unsupervised variational region-based segmentation [45, 33, 43, 38, 13, 4] rely on the level-set methodology [29, 23, 41, 34] which has been established as an elegant and efficient mathematical tool for the solution of problems involving evolving interfaces, offering robustness and tractability.

A significant precursor of recent work has been the Region Competition [46] method, which has helped clarify and unify different variational criteria and has introduced a probabilistic flavor in the curve evolution literature. The core idea of this algorithm is the maximization of the probability of the image observations I , using a set of M regions R_i , within which the observations are assumed to follow a simple region-specific parametric distribution $P(\cdot; a_i)$; an additional term on the length of the region borders, L_i is used to give rise to the following functional:

$$J(\Gamma, \{a_i\}) = \sum_{i=1}^M \frac{\mu}{2} \int_{\Gamma_i} ds - \iint_{R_i} \log P(I; a_i) \quad (6)$$

We have omitted the penalty on the number of regions used in [46], since in our case this remains fixed throughout the segmentation process. Calculus of variations yields the evolution of the region borders as the motion along the direction that assigns pixel observations to the region that models them better while maintaining the borders smooth:

$$\frac{\partial \Gamma_i}{\partial t} = -\mu \kappa \mathcal{N} + \log \frac{P(I; a_i)}{P(I; a_j)} \mathcal{N} \quad (7)$$

where κ is the curvature and \mathcal{N} the outward normal of front Γ_i and j is the neighboring region that competes with i for the observations at the interface position. Parameter estimation for the distributions $P(\cdot; a_i)$ is performed in alternation with curve evolution yielding an adaptive unsupervised image segmentation scheme.

In [33] this evolution algorithm has been brought together with the level set methodology and combined with edge-based terms, giving rise to the Geodesic Active Regions (GAR) algorithm

$$\frac{\partial \Gamma_i}{\partial t} = \lambda \log \frac{P(I; a_i)}{P(I; a_j)} - (1 - \lambda) [g(P_c) \kappa \mathcal{N} + (\nabla g(P_c) \cdot \mathcal{N}) \mathcal{N}] \quad (8)$$

where $g(\cdot)$ a monotonically decreasing function, P_c the probability of a pixel belonging to a boundary and λ determines the relative weights assigned to region- and edge-based information. The last term is inspired from the Geodesic Active Contour Model [16, 5] and forces the region borders to stay close to the locations where an edge detector responds strongly. Other successful algorithms like [43, 45] can be seen to be of the same essence with the original Region Competition algorithm; for example the model of [43] is based on the cartoon approximation to the Mumford-Shah functional [28], which in turn is a special case of the Region Competition functional for Gaussian distributions with equal variances.

2.4 Texture Features for Unsupervised Variational Segmentation

As mentioned earlier, even though the outputs of a filterbank may provide a rich description of the texture signal, their high dimensionality can lead to suboptimal segmentations. In the supervised texture segmentation scenario e.g. [32] this problem is bypassed by choosing the channels that maximally separate different textures. It is however harder to tackle the unsupervised problem, since choosing the best channels is equivalent to projecting the features onto a subspace where some *unknown a-priori* classes become maximally separated; this is usually performed using heuristic criteria, as e.g. in [39, 40].

In a recent attempt to alleviate the problem for unsupervised segmentation, Rousson et al. [38] have used a vector valued diffusion procedure to smooth a low-dimensional image descriptor, derived from local image derivatives. Combined with image intensity the resulting four-dimensional feature vector offers satisfactory results for the unsupervised segmentation of textured images. In the information-theoretic approach of [17] segmentation is accomplished without using a feature extraction stage, using as the sole criterion the maximization of the mutual information between region label and image intensity. In [13] distributions tuned to capture natural image statistics were shown to result in improved results when incorporated in curve evolution schemes for texture segmentation.

The representation of a texture in terms of its DCA features offers a low-dimensional feature vector, that is expressive enough for the discrimination of a wide variety of textures, encompassing information about texture strength, scale and orientation. Specifically, the feature vector we consider consists of the image intensity and the DCA components, namely amplitude, frequency magnitude and orientation. The distribution of the feature vector $P(\cdot; a_i)$ within region i is modelled as a product of a multivariate Gaussian for the first three dimensions and a von-Mises distribution for the orientation feature θ , which is analogous to the Gaussian distribution for orientational data:

$$P_{VM}(\theta; \theta_0, \kappa) \propto \exp(\kappa \cos(\theta - \theta_0)) \quad (9)$$

The parameters θ_0, κ of this distribution are estimated as in [9]. Segmentation results using the DCA feature vector can be found in Section V as well as in [19].

3 Regularized Demodulation

Let us consider an image I of K locally narrowband components $f_k(x, y)$ modelled by AM-FM signals and corrupted by a WSS zero-mean Gaussian noise field $w(x, y)$:

$$I(x, y) = \sum_{k=1}^K \underbrace{a_k(x, y) \cos(\omega_{k0} \cdot (x, y) + \phi_k(x, y))}_{\approx f_k(x, y)} + w(x, y)$$

For each component f_k , assuming a negligible AM-FM modeling error, its instantaneous frequency is given by $\omega_k = \omega_{k0} + \nabla \phi_k(x, y)$, where the carrier ω_{k0} is its mean frequency and $\phi_k(x, y)$ is the nonlinear phase part. The fundamental problem of demodulating the image I aims at estimating the instantaneous amplitudes $a_k(x, y)$ and frequencies $\omega_k(x, y)$. Unavoidable modeling errors of any demodulation algorithm, the presence of noise, interference from neighbor spectral components, and space discretization of the signal derivatives are possible sources that can cause errors in the demodulation of each narrowband component $f_k(x, y)$. Robustness in the AM-FM demodulation problem can be achieved in various ways, e.g. by optimizing any one or some of the following problems: (1) Reduction of the error in modeling each narrowband component $f_k(x, y)$ by a 2D AM-FM signal while maintaining some smoothness in the estimated amplitude and frequency modulation signals. (2) Suppression of noise. (3) Suppression of neighbor spectral components while estimating one component. (4) Regularization of derivatives. Simultaneously achieving all the above goals is a complex optimization task, which remains an unsolved problem. We present a regularized 2D energy operator and a related regularized 2D ESA that address some of the above problems in more than one combinations.

In the 1D case [30], given a narrowband signal $f(x)$ to model by $a(x) \cos(\omega_0 x + \phi(x))$, problem (1) has been given an optimum solution based on the Hilbert transform and analytic signal, which minimizes the mean-squared energy $\int |a'(x)|^2 dx$ of the amplitude derivative. This yields an optimum carrier frequency ω_0 as the center of gravity of the one-sided power spectrum $\int_0^\infty |F(\omega)|^2 d\omega$ of the given signal $f(x)$. In [36] it has been shown that, the ESA estimates of instantaneous amplitude and frequency of $f(x)$ bare a modeling error comparable to that of the Hilbert transform. However ESA has a better space-time resolution and much lower complexity. A simultaneous solution to problems (2) and (3) has been given in [3] by using a filterbank of bandpass filters, each centered at the spectral mean location of each narrowband component. The bandpass filtering increases the SNR and reduces the bias and variance of the ESA estimates of instantaneous amplitude and frequencies.

Given discrete image data, problem (4) immediately arises since the energy operator involves two differential operators. As analyzed in [35] for the problem of edge detection, two *regularized solutions*, which minimize the sum of the data approximation error and the energy of the second derivative of the approximating function, are (i) spline interpolation and (ii) convolution of the image data by a function that can be closely modelled by a Gaussian. In our problem which deals with narrowband but not necessarily lowpass signal components the Gaussian filter response must be modulated by a sine with carrier equal to the spectral mean location of the component. This yields a Gabor filter. In [8], the spline and the Gabor regularization of the energy operator and of the ESA were compared for 1D signals. This comparison yielded a slight superiority of the Gabor ESA.

Motivated by all the above, we propose a *2D Gabor ESA* algorithm for simultaneous filtering and demodulation. Let $I(x, y)$ be the continuous image, $g(x, y)$ the impulse of a real¹ 2D Gabor filter, and $f(x, y) = I(x, y) * g(x, y)$ the output of the Gabor filter. Since, convolution commutes with differentiation, the continuous 2D energy operator combined with Gabor bandpass filtering becomes

$$\Phi(f) = \Phi(I * g) = \|I * \nabla g\|^2 - (I * g)(I * \nabla^2 g) \quad (10)$$

Thus, the differential operators have been replaced by derivatives of the Gabor filter. The final algorithm for the Gabor energy operator (EO) becomes: (1) Find analytically and store all required differential formulae of

¹ If we use a complex Gabor filter, then we can use a 2D energy operator for complex-valued signals f defined in [24] by $C(f) = \Phi[\Re(f)] + \Phi[\Im(f)]$, i.e. the sum of the real energy operator applied to the real and imaginary part of the complex signal.

the Gabor function $g(x, y)$ evaluated on the pixel locations of a sampling grid $(i, j) = (i\Delta x, j\Delta y)$. We need three differential formulae: $g_x, g_y, \nabla^2 g$. (2) For estimating the instantaneous energy at the pixel locations (i, j) use the formula (10) of the combined continuous energy operator and Gabor filtering by using for each convolution the discrete convolution of the given image data $I[i, j]$ and the required Gabor derivative sampled at (i, j) .

Similarly, for 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 EO to compute the instantaneous energies of three image functions: $\Phi(f), \Phi(f_x = I * g_x) = \|I * \nabla g_x\|^2 - (I * g_x)(I * \nabla^2 g_x)$ and $\Phi(f_y = I * g_y)$. For all three energies we need seven Gabor differential formulae: $g_x, g_y, g_{xx}, g_{yy}, g_{xy}, \nabla^2 g_x, \nabla^2 g_y$. (2) Use the evaluated energies in the formula of the 2D continuous ESA. The 2D Gabor EO is computationally more intensive than the corresponding discrete EO, since it needs three convolutions (compared with one for the discrete case), but adds robustness and improved performance.

One approach to reduce the total complexity of applying the Gabor EO or the ESA to all filter outputs is to use the following modified procedure: (1) Apply Gabor bandpass filters to obtain all narrowband components $f_k = I * g_k$. (2) To each component $f_k(x, y)$ apply the following *Regularized Energy Operator* (REO)

$$\Phi_\sigma(f_k) = \|f_k * \nabla G_\sigma\|^2 - f_k(f_k * \nabla^2 G_\sigma)$$

where $G_\sigma(x, y) = (1/2\pi\sigma^2) \exp[-(x^2 + y^2)/2\sigma^2]$ is an isotropic Gaussian at regularization scale σ , and ∇G_σ and $\nabla^2 G_\sigma$ are the well-known gradient-of-Gaussian and Laplacian-of-Gaussian operators. The REO needs only 3 convolutions of the narrowband image component f_k with the $\partial/\partial x, \partial/\partial y$ and ∇^2 of the Gaussian. For the corresponding regularized ESA, we need the three regularized energies $\Phi_\sigma(f_k), \Phi_\sigma(\partial f_k/\partial x) = \|f_k * \nabla(\partial_x G_\sigma)\|^2 - f_k[f_k * \nabla^2(\partial_x G_\sigma)]$ and $\Phi_\sigma(\partial f_k/\partial y)$. All the Gaussian differential formulae are common for all filters and need to be computed once.

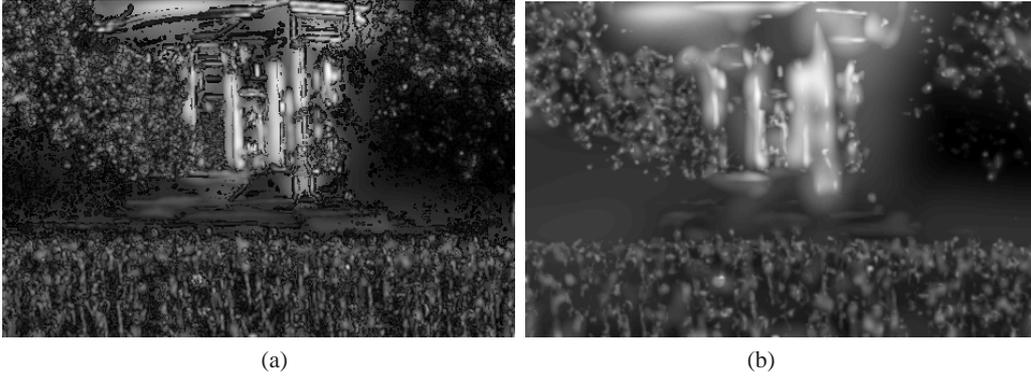


Fig. 1. Regularized Features: Dominant Amplitude (a) ESA (b) Gabor ESA.

4 Probabilistic Cue Integration

A problem faced by our algorithm, as well as most feature-based segmentation algorithms, is that at areas where the model underlying the feature extraction process fails to accurately capture signal behavior the features may be meaningless and can drive the segmentation to suboptimal solutions. For example, in the presence of edges the DCA-based amplitude is typically high, describing large oscillations while at smooth regions the orientation of the frequency vector varies erratically. In this section we introduce a method that renders the segmentation process immune to such problems by automatically choosing which features the evolution equations should rely upon.

Our method uses a confidence measure assigned to the features used to drive the segmentation process; this may be hard to obtain generally, but can be naturally accomplished in our case where a specific *generative* model can be devised, i.e. a model that for a specific set of parameters offers a prediction and a related

likelihood expression for its observations. When multiple models are used for feature extraction, a posterior distribution can be defined using these likelihood expressions based on Bayes' rule, thereby indicating which of these most accurately captures the observations.

Below we briefly review our contributions [18, 20] in providing such a framework for the DCA model and subsequently provide a modification of the Region Competition/GAR algorithm that allows for the fusion of the extracted features in a simple and efficient manner.

4.1 Generative Models for Features

The generative model we introduce for texture explains the image neighborhood around a specific point in terms of two sub-models, namely a texture model and a generic background model; the degree to which these two sub-models contribute to the explanation of an observation depends on its distance from the point around which the model is defined. In all of the following the treatment will consider the signals as one-dimensional, for the sake of notational clarity. The synthesized prediction $S^T(x)$ at point x of the texture sub-model defined around point 0 can be expressed as the sum of a sinusoidal of fixed frequency ω and phase offset ϕ and a DC component B : $S^T(x) = A \cos(\omega x + \phi) + B$, while the background sub-model is considered to be a uniform distribution. The spatial variation of our confidence in these two sub-models can be phrased as:

$$P(I(x)|x, 0) = G(x)P(I(x)|S^T(x)) + (1 - G(x))c, \quad (11)$$

where $P(I(x)|x, 0)$ is the likelihood of the observation at point x , given the model related to the texture hypothesis that is defined around point 0; $G(x)$ accounts for the locality of the modeling process and is taken to be a Gaussian function normalized so that $G(0) = 1$. The corruption of the texture synthesis $S^T(x)$ is modelled by $P(I(x)|S^T(x))$ which is a Gaussian distribution with mean $S^T(x)$ and unknown, constant variance while c is the constant term contributed by the uniform background model. Using the quadrature pair $g_e = \cos(\omega x)G(x)$ and $g_o = \sin(\omega x)G(x)$ of Gabor filters centered around frequency ω , in [18, 20] it is shown that the quantity $A_k = \sqrt{(g_e * I)^2 + (g_o * I)^2}$ is related to the likelihood of the image observations under the probabilistic model outlined above. In this sense, picking the channel with the largest amplitude estimate amounts to choosing the channel that explains the data best.

To apply the same rationale to edge detection, we used the observation that edges are *phase congruent* signals [27], i.e. signals of the form $A \sum_k a_k \cos(\omega k + \phi)$; for $\phi = 0$ we derive line (triangular) edges and for $\phi = \frac{\pi}{2}$ step edges. Again, using an appropriate quadrature filter pair of odd and even filters we can derive an amplitude estimate related to the likelihood of the observations under this model [20]. A model that complements these two classes is the smooth signal model, for which we use the locally DC signal, using again an expression of the form Eq. (11) to account for the locality of the decision made.

Using these three models we are able to estimate the probability of a pixel having been generated by one of the three hypotheses considered based on Bayes' rule. Applying this approach on natural images gives visually appealing results as shown in Fig. 2, since textured areas are correctly discriminated from edges, matching closely what a human would call a texture or an edge. Given that the previously described models do not take into account higher order structure or grouping cues, the results are quite satisfactory. A more extensive presentation of the approach outlined above can be found in [18, 20].

4.2 Cue Integration for Region Competition

In this section we present a modification of the Region Competition algorithm that can exploit the confidence measures assigned to the feature set used to drive the segmentation process. This fits naturally with the models used for feature extraction in our work, but its application is not constrained to these.

The rationale underlying the Region Competition evolution equations is to assign to each region the observations that it can most accurately explain in terms of its distribution. This is implemented using a probabilistic balloon force [6], that pushes the front of region i along its outward normal with a force proportional to:

$$\mathcal{O}_I = \log \frac{P_i(I; a_i)}{P_j(I; a_j)},$$

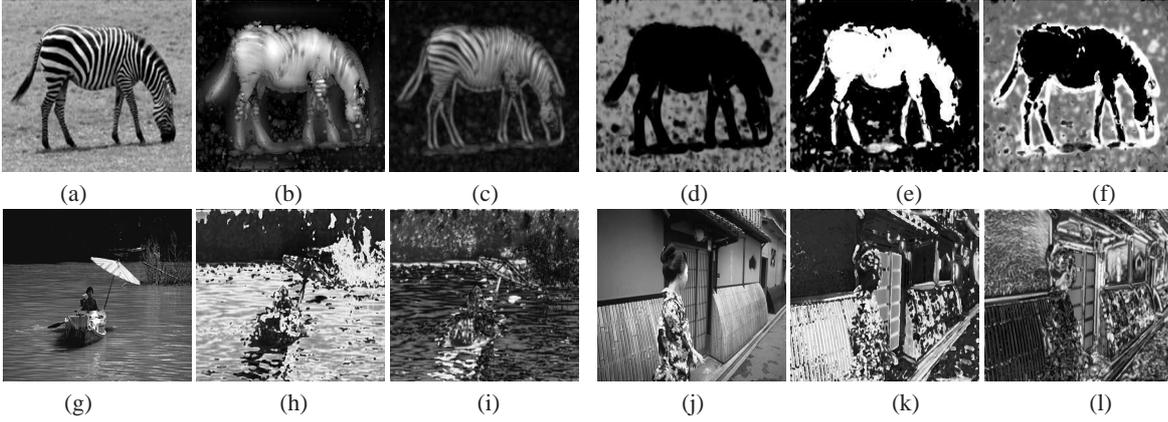


Fig. 2. From [20]: Textured areas can be successfully discriminated from edges, using simple generative models. Top row: (a) Input image, (b) texture and (c) edge model amplitude estimates, respectively and model-based probabilities of (d) smooth, (e) textured and (f) edge regions respectively. Bottom row: (g)/(j) Input images and posterior probabilities of (h)/(k) texture and (i)/(l) edge models respectively.

where j is the competing neighboring region. This quantity, termed *log-odds* quantifies the degree to which the observation I is more likely under hypothesis i than under hypothesis j . Supra - Bayesian fusion methods [14] consider this quantity \mathcal{O}_I as a random variable, which follows a Gaussian distribution conditioned on the actual class of the data I :

$$P(\mathcal{O}_I|i) \propto N(\mu_i, \sigma^2), \quad P(\mathcal{O}_I|j) \propto N(\mu_j, \sigma^2)$$

This quantifies the certainty associated with any decision made based on \mathcal{O}_I : a large σ , i.e. a low confidence in log-odd accuracy can diminish the effect of a large value of \mathcal{O}_I . From a good classifier we would generally expect that $\mu_i \gg \mu_j$ with a low σ , which means that if the data I are due to hypothesis i , then it is very probable that \mathcal{O}_I will take a high value and vice versa.

Based on the above approach, the results of N classifiers using different features or different classification methods can be easily integrated [14]: their log-odds $\mathcal{O} = [\mathcal{O}_1, \dots, \mathcal{O}_N]$ are viewed as a multidimensional random variable that follows a Gaussian distribution conditioned on the class of the data i.e.

$$P(\mathcal{O}|i) \propto N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \quad P(\mathcal{O}|j) \propto N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$$

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

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

We thereby build a decision using as features the outputs of classifiers, instead of the actual features. For the special case where the classifier outputs are uncorrelated, 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 the $\boldsymbol{\mu}_i$ vector equal unity, so that we have

$$\log \frac{P(i|\mathcal{O})}{P(j|\mathcal{O})} = \sum_c \frac{\mathcal{O}_c}{\sigma_c^2}$$

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 his decision and vice versa. The use of log-odds is particularly convenient, in that we express this weighting operation in terms of a summation.

In our case, we consider that the assignment of an observation to region i or region j is a decision taken by fusing the assessments of two experts, where the decision of the first is based on the texture features and that of the second on the intensity values. Using the previous notation we have

$$\mathcal{O}_T = \log \frac{P^T(F_T; a_i^T)}{P^T(F_T; a_j^T)}, \quad \mathcal{O}_S = \log \frac{P^S(F_S; a_i^S)}{P^S(F_S; a_j^S)}$$

where by $P^M(F_M; a_i^M)$ we denote the likelihood of the feature-set F_M extracted based on hypothesis M (texture-T or smooth-S) under the hypothesis-specific distribution P^M of region i , whose parameters are a_i^T . For the final decision, each expert's opinion is weighted by the probability of each hypothesis, estimated as described in the previous subsection; 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 6 thus becomes:

$$\frac{\partial \Gamma_i}{\partial t} = \left[\sum_{c \in T, S} w_c \log \frac{P^c(I_c; \alpha_i)}{P^c(I_c; \alpha_j)} - \mu \kappa \right] \mathcal{N} \quad (12)$$

where the index c ranges over the cues that are being fused and w_c is the related cue weight. Further, as in [33] one can introduce edge based information, but now exploiting the model-based probability of edge in which case the evolution equation becomes:

$$\frac{\partial \Gamma_i}{\partial t} = \left[\sum_{c \in T, S} w_c \log \frac{P^c(I_c; \alpha_i)}{P^c(I_c; \alpha_j)} \right] \mathcal{N} - w_e [g \kappa \mathcal{N} + (\nabla g \cdot \mathcal{N}) \mathcal{N}] \quad (13)$$

with w_e being the probability of the edge hypothesis. This does not interfere with the edge detection procedure used to estimate ∇g , since it simply provides an indication of how important the edge information should be deemed.

We note here that the weights entering the fusion equations do not have to be estimated using the specific models described previously; as an alternative that we intend to explore in future work, the celebrated U+V decomposition [44] could be used to indicate regions with a strong texture component, and subsequently provide weights for a fusion algorithm.

5 Experimental Evaluation

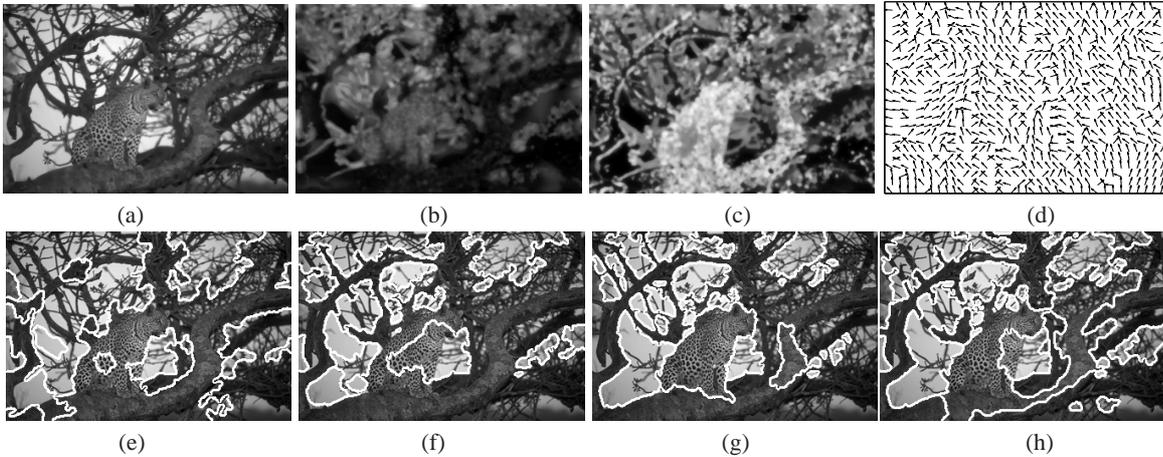


Fig. 3. Top Row: Features used for Segmentation. (a) Intensity, (b) Amplitude, (c) Freq. Magnitude, (d) Freq. Orientation vectors, Bottom Row: Segmentation results, using (e) the original evolution equations [46, 32] as in [19], (f) the fusion evolution equations, excluding the orientation channel and (g) including the orientation channel, (h) Diffusion-based features [38] and the original evolution equations. Please see text for details.

As described in [19], curve evolution is implemented using level-set methods [41, 42] along the lines of [33]. As in [40, 38, 45] the distribution of the data inside each region is learned in parallel with the evolution process, resulting in an adaptive scheme. For all the results presented in this work the regions have been initialized so as to partition the whole image in interleaved thin parallel strips, while we have observed that the results do not depend substantially on the initialization.

In Fig. 3 (a)-(d) we present the modulation features extracted via DCA and segmentation results using four alternative schemes: In Fig. 3(e) the results of curve evolution along the lines of [19] are shown, using the 3-dimensional DCA-based texture descriptor while in (f) we show results using the cue integration scheme described in this work. In (g) the orientation channel is added to the feature vector, with no performance degradation on smooth areas, contrary to (e), and improved boundary localization. In the fused cases, (f) and (g), the region boundaries tend to accurately capture the object borders, while in (e) the erroneously estimated texture features prevent them from doing so. Further results are provided in Fig. 4, where we generally observe that the fused features give better segmentations results, with the region borders accurately locating the object borders. The effect of the orientation features which behave erratically at smooth areas is diminished, due to the smaller weight assigned to the texture hypothesis at these regions.

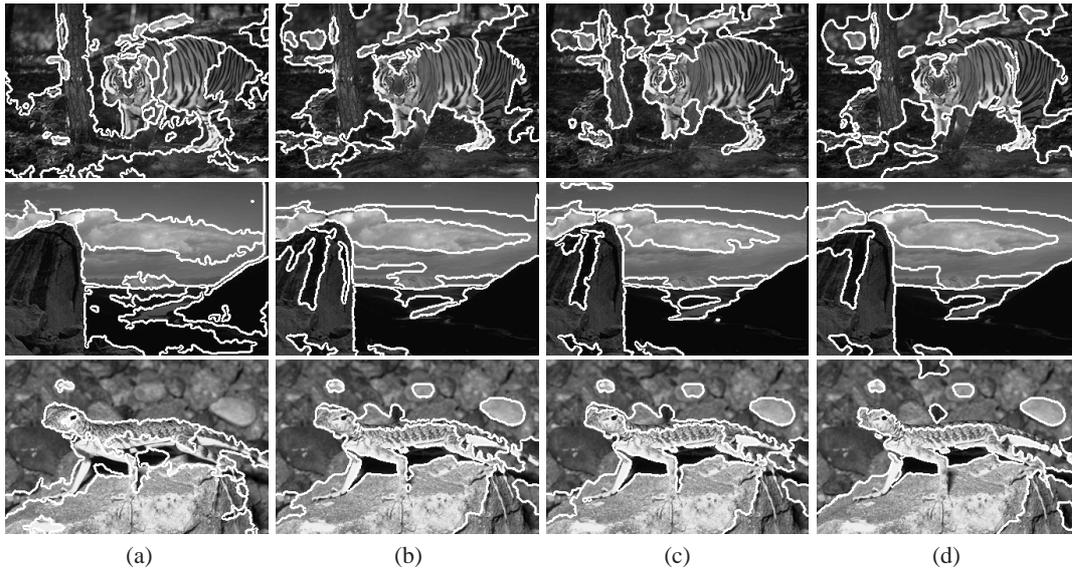


Fig. 4. Segmentation Comparisons: (a) original evolution equations, (b) the fusion evolution equations, excluding the orientation channel and (c) including the orientation channel, (d) Diffusion-based features [38].

In order to obtain quantitative results, segmentation results for an increasing numbers of fronts were derived for the whole Berkeley test set and compared to human segmentations based on the Bidirectional Consistency Error (BCE) measure introduced in [25]. This measure quantifies in a smooth manner the overlap between a machine generated segmentation and a set of manual segmentations, and is minimized when for every machine-generated segment there is at least one human-generated segmentation wherein the segment is contained as a whole and vice versa.

Initially we compared the performance of the raw AM-FM features used in [19] to that of the nonlinear diffusion-based feature set [38]; in all the related results presented herein some uncertainty is retained due to potential inaccuracies in the implementation of the algorithms in [38]. For this comparison, the orientation features were omitted, in order to avoid spoiling the feature vector at smooth regions. Comparing the histograms in Fig.s. 5(a)(b) the BCE can be seen to be lower for the AM-FM features, indicating their potential to accurately describe a textured region in terms of a low-dimensional feature vector. As already mentioned, the introduction of the orientation channel in the feature vector typically results in segmentation performance degradation in the cases of smoothly varying areas and edges; a point that we make in this paper is that using a fusion algorithm counteracts these defects and allows the orientation channel to be incorporated in a manner that is immune to its behavior on smooth regions. This can be seen by a comparison between Fig.s 5(b)(c), where the BCE is seen to have almost the same distribution; that the introduction of the orientation vector does not lead to a systematic improvement in the results can be due the fact that the data set provided in [26] does not contain many heavily textured images, where the information carried by the orientation channel is most valuable.

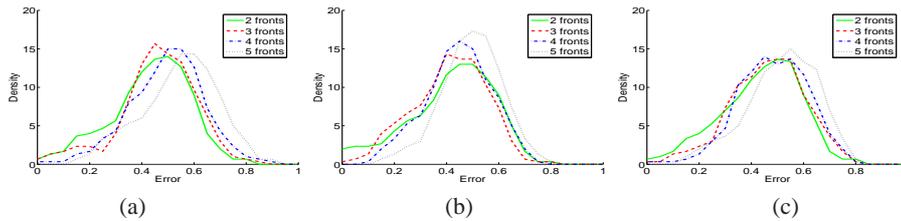


Fig. 5. Berkeley Benchmark evaluation: Histogram of Bidirectional Consistency Error [25] over 100 test images for varying number of fronts (2-5) for: **(a)** Diffusion-based features **(b)** Unfused data, without orientation **(c)** Fused data with orientation. Please see text for details.

6 Conclusions

Multicomponent AM-FM models propose a powerful approach to the representation, analysis and segmentation of textured images. Our contributions presented herein lie in (1) the introduction of a regularized demodulation algorithm that can alleviate discretization problems and introduce increased noise robustness and (2) the probabilistic integration of features related to different image models, based on a modification of the region competition evolution equations. Systematic comparisons have demonstrated that the derived features compare favorably to those used by the current state-of-the-art methods, indicating the appropriateness of modulation features for unsupervised textured image segmentation.

References

1. BOVIK, A., CLARK, M., AND GEISLER, W. Multichannel Texture Analysis using Localized Spatial Filters. *IEEE Trans. PAMI* 12, 1 (1990), 55–73.
2. BOVIK, A. C., GOPAL, N., EMMOTH, T., AND RESTREPO, A. Localized Measurement of Emergent Image Frequencies by Gabor Wavelets. *IEEE Trans. Information Theory* 38 (1992), 691–712.
3. BOVIK, A. C., MARAGOS, P., AND QUATIERI, T. F. AM-FM Energy Detection and Separation in Noise Using Multiband Energy Operators. *IEEE Trans. Signal Processing* 41 (1993), 3245–3265.
4. BROX, T., AND WEICKERT, J. A TV Flow based Local Scale Measure for Texture Discrimination. In *ECCV* (2004).
5. CASELLES, V., KIMMEL, R., AND SAPIRO, G. Geodesic Active Contours. *Int.l. J. of Comp. Vision* 22, 1 (1997), 61–79.
6. COHEN, L. On Active Contour Models and Balloons. *Computer Vision, Graphics, and Image Processing: Image Understanding* 53, 2 (1991), 211–218.
7. DAUGMAN, J. Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two-Dimensional Visual Cortical Filters. *Journal of the Optical Society of America* 2, 7 (1985), 160–169.
8. DIMITRIADIS, D., AND MARAGOS, P. Robust Energy Demodulation Based on Continuous Models with Application to Speech Recognition. In *Eurospeech* (2003).
9. GUMBEL, E., GREENWOOD, A., AND DURAND, D. The Circular Normal Distribution: Theory and Tables. *Journal of the American Statistical Association* 48, 261 (1953), 131–152.
10. HAVLICEK, J., AND BOVIK, A. Image Modulation Models. In *Handbook of Image and Video Processing*, A. Bovik, Ed. 2000, pp. 305–316.
11. HAVLICEK, J., HARDING, D., AND BOVIK, A. The Multi-Component AM-FM Image Representation. *IEEE Trans. Im. Proc.* 5, 6 (1996), 1094–1100.
12. HAVLICEK, J., HARDING, D., AND BOVIK, A. Multidimensional Quasi-Eigenfunction Approximations and Multicomponent AM-FM Models. *IEEE Trans. Im. Proc.* 9, 2 (2000), 227–242.
13. HEILER, M., AND SCHNORR, C. Natural Image Statistics for Natural Image Segmentation. *Int.l. J. of Comp. Vision* 63, 1 (2005), 5–19.
14. JACOBS, R. Methods for Combining Experts' Probability Assesements. *Neural Computation*, 7 (1995), 867–888.
15. KASS, M., WITKIN, A., AND TERZOPOULOS, D. Snakes: Active Contour Models. In *ICCV* (1987).
16. KICHENASSAMY, S., KUMAR, A., OLVER, P., TANNENBAUM, A., AND YEZZI, A. Gradient Flows and Geometric Active Contour Models. In *ICCV* (1995).
17. KIM, J., FISHER, J., YEZZI, A., CETIN, M., AND WILLSKY, A. Nonparametric Methods for Image Segmentation using Information Theory and Curve Evolution. In *Int.l Conf. on Image Processing* (2002).
18. KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Advances in Texture Analysis: Energy Dominant Components and Multiple Hypothesis Testing. In *Int.l Conf. on Image Processing* (2004).

19. KOKKINOS, I., EVANGELOPOULOS, G., AND MARAGOS, P. Modulation-Feature based Textured Image Segmentation Using Curve Evolution. In *Int.l Conf. on Image Processing* (2004).
20. KOKKINOS, I., AND MARAGOS, P. A Detection-Theoretic Approach to Texture and Edge Discrimination. In *4th International Workshop on Texture Analysis and Synthesis* (2005).
21. LEE, T. S., MUMFORD, D., AND YUILLE, A. Texture Segmentation by Minimizing Vector Valued Energy Functionals. In *ECCV* (1992).
22. MALIK, J., AND PERONA, P. Preatentive Texture Discrimination with Early Vision Mechanisms. *J. Opt. Soc. Amer. (A)* 7, 5 (1990), 923–932.
23. MALLADI, R., SETHIAN, J., AND VEMURI, B. Shape Modeling with Front Propagation: A Level-Set Approach. *IEEE Trans. PAMI* 17, 2 (1995), 158–175.
24. MARAGOS, P., AND BOVIK, A. Image Demodulation Using Multidimensional Energy Separation. *Journal of the Optical Society of America* 12, 9 (1995), 1867–1876.
25. MARTIN, D. *An Empirical Approach to Grouping and Segmentation*. PhD thesis, Univ. California at Berkeley, 2002.
26. MARTIN, D., FOWLKES, C., TAL, D., AND MALIK, J. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In *ICCV* (2001).
27. MORRONE, C., AND BURR, D. Feature Detection in Human Vision: a Phase-Dependent Energy Model. *Proceedings of the Royal Society of London B* 235 (1988), 221–245.
28. MUMFORD, D., AND SHAH, J. Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics* 42, 5 (1989), 577–685.
29. OSHER, S., AND SETHIAN, J. Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *Journal of Computational Physics* 79 (1988), 12–49.
30. PAPOULIS, A. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1984.
31. PARAGIOS, N., AND DERICHE, R. Geodesic Active Contours for Supervised Texture Segmentation. In *CVPR* (1999).
32. PARAGIOS, N., AND DERICHE, R. Geodesic Active Regions: A new Paradigm to Deal with Frame Partition Problems in Computer Vision. *Journal of Visual Communication and Image Representation* 13, 1 (2002), 249–268.
33. PARAGIOS, N., AND DERICHE, R. Geodesic Active Regions and Level Set Methods for Supervised Texture Segmentation. *Int.l. J. of Comp. Vision* 46, 3 (2002), 223–247.
34. PARAGIOS, N., AND OSHER, S. *Geometric Level Set Methods in Imaging, Vision & Graphics*. 2002.
35. POGGIO, T. A., VOORHEES, H., AND YUILLE, A. A Regularized Solution to Edge Detection. *J. Complexity* 4, 2 (1988), 106–128.
36. POTAMIANOS, A., AND MARAGOS, P. A Comparison of the Energy Operator and Hilbert Transform Approaches for Signal and Speech Demodulation. *Signal Processing* 37, 1 (1994), 95–120.
37. RAY, N., HAVLICEK, J., ACTON, S., AND PATTICHIS, M. Active Contour Segmentation Guided by AM-FM DCA. In *Int.l Conf. on Image Processing* (2001).
38. ROUSSON, M., BROX, T., AND DERICHE, R. Active Unsupervised Texture Segmentation on a Diffusion-Based Feature Space. In *CVPR* (2003).
39. SAGIV, C., SOCHEN, N. A., AND ZEEVI, Y. Y. Texture Segmentation via a Diffusion-Segmentation Scheme in the Gabor Feature Space. In *2nd International Workshop on Texture Analysis and Synthesis* (2002).
40. SANDBERG, B., CHAN, T., AND VESE, L. A Level-Set and Gabor-based Active Contour Algorithm for Segmenting Textured Images. Tech. Rep. 02-39, UCLA CAM, 2002.
41. SETHIAN, J. *Level Set Methods*. Cambridge University Press, 1996.
42. SETHIAN, J. A. A Fast Marching Method for Monotonically Advancing Fronts. *Proc. Nat.l Academy of Sciences of USA* 93, 4 (1996), 1591–1595.
43. VESE, L., AND CHAN, T. Active Contours Without Edges. *IEEE Trans. Im. Proc.* 10, 2 (2001), 266–277.
44. VESE, L., AND OSHER, S. J. Modeling Textures with Total Variation Minimization and Oscillating Patterns in Image Processing. *Journal of Scientific Computing* 19, 1-3 (2003), 553–572.
45. YEZZI, A., TSAI, A., AND WILLSKY, A. A Statistical Approach to Snakes for Bimodal and Trimodal Imagery. In *ICCV* (1999).
46. ZHU, S., AND YUILLE, A. Region Competition: Unifying Snakes, Region Growing and Bayes/MDL for Multiband Image Segmentation. *IEEE Trans. PAMI* 18, 9 (1996), 884–900.