

Generalized Flooding and Multicue PDE-Based Image Segmentation

Anastasia Sofou and Petros Maragos, *Fellow, IEEE*

Abstract—Image segmentation remains an important, but hard-to-solve, problem since it appears to be application dependent with usually no *a priori* information available regarding the image structure. Moreover, the increasing demands of image analysis tasks in terms of segmentation results' quality introduce the necessity of employing multiple cues for improving image segmentation results. In this paper, we attempt to incorporate cues such as intensity contrast, region size, and texture in the segmentation procedure and derive improved results compared to using individual cues separately. We emphasize on the overall segmentation procedure, and we propose efficient simplification operators and feature extraction schemes, capable of quantifying important characteristics, like geometrical complexity, rate of change in local contrast variations, and orientation, that eventually favor the final segmentation result. Based on the well-known morphological paradigm of watershed transform segmentation, which exploits intensity contrast and region size criteria, we investigate its partial differential equation (PDE) formulation, and we extend it in order to satisfy various flooding criteria, thus making it applicable to a wider range of images. Going a step further, we introduce a segmentation scheme that couples contrast criteria in flooding with texture information. The modeling of the proposed scheme is done via PDEs and the efficient incorporation of the available contrast and texture information, is done by selecting an appropriate cartoon-texture image decomposition scheme. The proposed coupled segmentation scheme is driven by two separate image components: cartoon U (for contrast information) and texture component V . The performance of the proposed segmentation scheme is demonstrated through a complete set of experimental results and substantiated using quantitative and qualitative criteria.

Index Terms—Feature extraction, morphological filtering, partial differential equation (PDE), segmentation, topographic flooding, $U + V$ image decomposition, watershed.

I. INTRODUCTION

IMAGE segmentation is one of the most important, yet complicated, problems in the field of computer vision. Its critical role is derived by the fact that it forms the basis for most subsequent image analysis tasks. It is commonly accepted that

the term *segmentation* covers a wider range of image processing tasks than the final partitioning of the image plane into disjoint regions. In this paper, we treat the segmentation problem as a set of procedures that need to be followed starting from the initial image and yielding the final partitioning perceived either as a region map or a segmentation boundary. Thus, the task of segmentation, independently of the method used to achieve the partitioning, can be divided into the following stages: i) *image preprocessing*, ii) *feature extraction*, and iii) *partitioning* into disjoint regions. The preprocessing stage encompasses a wide range of subtasks such as image simplification (enhancement, smoothing, noise reduction, redundant information removal) resulting in an image consisting mostly of flat and large regions, as well as image decomposition into constituent parts. The feature extraction deals with gradient features computation, texture measurements, markers extraction (small homogeneous regions), whereas the final stage of partitioning is the application of the selected segmentation algorithm so as to produce a region map of the image. Optionally, there can be a subsequent stage, namely *postprocessing*, where the segmented image can be further processed according to predefined criteria relevant to region properties, possibly merge or suppress undesired regions so as to lead to the final result.

Many different segmentation methodologies have been proposed and, depending on their approach to the problem, can be characterized either as boundary-based methods relying on the information provided by the object/region boundaries, or region-based exploiting information provided by the entire regions (contrast, texture properties, etc). A well-known region-based segmentation methodology, which has attracted the interest of many researchers for years, is the *watershed transform* [1]–[3], initially proposed as the solution to segmentation in the field of mathematical morphology. Throughout the years, it has been established as a very robust and efficient segmentation method, applicable to many different cases, especially when coupled with other nonlinear multiscale morphological operators. Compared to other methodologies, watershed has several advantages, including proper handling of gaps and placement of boundaries at most significant edges. From the morphological point of view, it is the process according to which the image is considered as a topographic relief and flooded uniformly from predefined sources (such as regional minima or selected markers). Aside from its implementations either via immersion simulations [2] or hierarchical queues [3] using an ordering relation in flooding, the watershed transform has also been modeled in a continuous way employing the eikonal partial differential equation (PDE) [4]–[6]. There are two categories of numerical implementations that can be used to solve

Manuscript received February 11, 2007; revised November 21, 2007. This work was supported in part by the Greek Secretariat for Research and Technology under the program IIENE Δ – 2001 and in part by the European Network of Excellence MUSCLE and the EU-GR program PYTHAGORAS. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Giovanni Poggi.

The authors are with the School of Electrical and Computer Engineering, National Technical University of Athens, Athens 15773, Greece (e-mail: sofou@cs.ntua.gr; maragos@cs.ntua.gr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2007.916156

the eikonal equation: i) implementations based on discrete gray-weighted distance transform (GWDT) with chamfer metrics [5] and ii) implementations based on PDEs modeling curve evolution with level sets [7], [6], [8]. The rationale for using PDEs lies in better and more intuitive mathematical modeling, connections with physics, and better approximation to the continuous geometry of the problem. Besides the efforts made to investigate the watershed transform and formulate it using different tools than the ones provided by mathematical morphology, there have been many attempts in the direction of integrating watershed with other segmentation approaches, such as watersnakes [9], where a connection between the watershed and minimization of an energy function is established.

Motivated by the efficacy of watershed transform along with latest trends in image segmentation research that encourage combination of different cues [10]–[12], we try to incorporate the generalized flooding concept of watershed, thus exploiting intensity contrast and region size criteria [13] with other perceptually meaningful image characteristics, such as texture, aiming at improved segmentation. Additionally, we aim at integrating the aforementioned ideas with PDE modeling. Preliminary results of this effort have been shown in [14], with applications in specialized data of soilsection images in [15]. Moreover, in order to strengthen the segmentation efficiency, we treat the segmentation problem as set of subtasks, emphasizing particularly on every segmentation stage, utilizing the most appropriate set of tools for the image processing tasks, considering also the fact that the quality of each stage's output affects the overall segmentation result.

In this paper, we propose well-motivated and efficient image simplification, texture modeling, image decomposition, and marker extraction techniques as necessary tasks of the presegmentation part of image analysis. Regarding the stage of partitioning algorithm application, we focus on watershed-based segmentation techniques. Specifically, we investigate the PDE watershed formulation and extend it in order to satisfy other flooding criteria, such as region size and volume, thus making it applicable to a wider range of images. Additionally, we incorporate different cues in watershed flooding by selecting a leveling-based $U + V$ image decomposition as an efficient way to separate image geometry from texture. We propose a watershed-like segmentation scheme that couples contrast, size, and texture information driven by two separate image components: cartoon U (for contrast information) and texture V . The modeling of the proposed scheme is done via PDEs using ideas from curve evolution and level sets. The implementation is done by using and adapting specialized level set methodologies, thus ensuring speed and reduced computational cost. Through a complete set of experimental results, we demonstrate the performance of the proposed segmentation scheme. Furthermore, we provide quantitative and qualitative criteria that substantiate the way the proposed methodology outperforms the established watershed transform.

The main contributions of the presented research work can be summarized as follows:

- 1) PDE-based formulation and extension of flooding methods for generalized watershed type segmentation;
- 2) PDE-based formulation of uniform volume flooding;
- 3) PDE geometry and texture driven segmentation;
- 4) extension of fast marching level-set-based method (FMM) to accommodate the implementation of the above PDEs;
- 5) multiple region PDE level-set segmentation;
- 6) integrated segmentation system that combines 1) and 2), plus state-of-the-art multiscale methods for preprocessing and feature/marker extraction;
- 7) improved evaluation criteria used over a large variety of images.

The rest of the paper is organized as follows. Section II provides the methodologies used for image simplification, decomposition and feature extraction, thus preparing the image for the application of segmentation algorithm. In Section III, we investigate the generalized watershed implemented via PDEs and extend it to incorporate varying flooding criteria, whereas, in Section IV, we diversify the previously presented ideas to embody contrast, size and texture criteria in a single segmentation algorithm. In Section V, we provide mature experiments as well as quality criteria and comparative results.

II. IMAGE SIMPLIFICATION, DECOMPOSITION, AND FEATURE EXTRACTION

It is almost impossible to avoid the presence of noise in an image usually introduced in the formation, acquisition, transmission process etc, and visualized either as distortion, artifacts or meaningless structures. Images may also contain texture in the form of repeated patterns of small scale details. Depending on what the human visual system can classify as salient areas or meaningful objects, not all image structures are of the same importance. Therefore, an image needs elaborate processing so that its structure is simplified but at the same time key features are accurately preserved. Additionally, it can be further processed and decomposed to its constituent components, each of which encapsulates information about distinct prominent image characteristics, such as contrast, texture, etc. The aforementioned processing aids the feature extraction procedure as well as the final partitioning stage. Next, we provide a detailed discussion about each of the subtasks mentioned above focusing on the employed tools.

A. Image Simplification

The simplification stage is concerned with noise and redundant information removal, resulting in an image with smoother structure, easier to handle and more appropriate for further processing such as feature extraction and partitioning. The filtering the image has to undergo should retain meaningful information but at the same time suppress pointless structures without causing boundary blurring or contour displacement. An efficient family of filters that have the aforementioned properties are the morphological connected operators. These are region-based filtering tools that do not modify individual pixel values but directly act on connected components of the space where the image is constant, referred to as flat zones. Intuitively connected operators can merge flat zones by removing boundaries between them, but cannot create new boundaries nor shift existing ones, therefore having very good contour preservation properties. The related literature consists of theoretical studies [16]–[21], and algorithm developments [22]–[24]. For image

simplification, we use the following connected operators for contrast/area/volume filtering and levelings.

1) *Contrast Filtering—Reconstruction Operators*: The graylevel reconstruction opening of an image $I(x, y)$ given a marker signal $M \leq I$ is

$$\rho^-(M|I) = \lim_{n \rightarrow \infty} F_n, F_n = \delta_B(F_{n-1}|I), F_0 = M \quad (1)$$

where $\delta_B(M|I) = (M \oplus B) \wedge I$ is the conditional dilation of M by a unit disk B constrained by I . The above definition is valid both for continuous and discrete¹ images. Dually, the reconstruction closing of $I(x, y)$ given a marker $M \geq I$ is

$$\rho^+(M|I) = \lim_{n \rightarrow \infty} F_n, F_n = \varepsilon_B(F_{n-1}|I), F_0 = M \quad (2)$$

where $\varepsilon_B(M|I) = (M \ominus B) \vee I$ is the conditional erosion of M by B constrained by I . For contrast filtering, we set the marker $M = I - h$ and $M = I + h$ for reconstruction opening and closing, respectively. The parameter h is a constant determined empirically to match the image peaks/valleys range and controls the contrast of the bright/dark connected components that will be merged.

Generalized openings γ_i and closings φ_i are often combined sequentially at increasing scale $i = 1, 2, \dots, n$ to produce *alternating sequential filters (ASF)* [16]: $\Psi_{ASF}(I) = \varphi_n \gamma_n \dots \varphi_2 \gamma_2 \varphi_1 \gamma_1(I)$. Choosing reconstruction openings and closings yields an ASF that accomplishes dual reconstruction at sequential multiple scales.

2) *Self-Dual Filtering—Levelings*: The above reconstruction operators are asymmetric since they allow simplification of either bright or dark image components. Symmetric simplification of image components requires self-dual filters, such as the **levelings**, which are nonlinear, increasing, and idempotent filters with many interesting scale-space properties [19], [20]. Levelings treat symmetrically the image foreground and background; further, they can be analyzed as composition of reconstruction opening and closing. They were defined geometrically in [19] via the property that the variation of the leveling between two neighbor pixels is bracketed by a larger same-sign variation in the reference image I . In [21], they were defined algebraically as fixed points of iterated triphase operators that switch among three phases, an expansion, a contraction, and the reference I . They operate on a reference image I by locally expanding/shrinking an initial marker image M , and globally constraining the marker evolution by the reference image. In the simplest case, iterations of the image operator $\lambda(F|I) = (\delta(F) \wedge I) \vee \varepsilon(F)$, where $\delta(F)$ (resp., $\varepsilon(F)$) is a dilation (resp., erosion) of F by a small disk, yield in the limit the leveling of I w.r.t. M

$$\Lambda(M|I) = \lim_{k \rightarrow \infty} F_k, F_k = \lambda(F_{k-1}|I), F_0 = M. \quad (3)$$

¹For example, the conditional dilation of a discrete marker image $M[i, j]$ w.r.t. a reference image $I[i, j]$ is implemented as follows:

$$\delta_B(M|I)[i, j] = \min \left(I[i, j], \max_{(k, \ell) \in B} M[i - k, j - \ell] \right)$$

where B is the 5-pixel (9-pixel) diamond (square) for 4-type (8-type) foreground connectivity.

Levelings preserve the coupling and sense of variation in neighbor image values and do not create any new regional maxima or minima across scales. In practice, they can reconstruct whole image objects with exact preservation of their boundaries and edges. In this reconstruction process, they simplify the original image by completely eliminating smaller objects inside which the marker cannot fit.

3) *Area Filtering—Area Operators*: The area opening (closing) of size n keeps only the light (dark) connected components whose area (number of pixels) is equal or greater than a threshold n . For a binary image-set X , its area opening is defined by $\alpha_n^-(X) = \bigcup \{X_j : \text{Area}(X_j) \geq n\}$ where $X = \bigcup_i X_i$ is the disjoint union of connected components X_i . By complementation, the binary area closing is defined as $\alpha_n^+(X) = [\alpha_n^-(X^c)]^c$. The grayscale area opening and closing can be defined based on the binary area opening and closing definition via threshold superposition

$$\alpha_n^-(I)(x, y) = \sup \{h : (x, y) \in \alpha_n^-(T_h(I))\} \quad (4)$$

$$\alpha_n^+(I)(x, y) = \sup \{h : (x, y) \in \alpha_n^+(T_h(I))\} \quad (5)$$

where $T_h(I)$ are the upper level sets of the image I by thresholding it at level h

$$T_h(I) = \{(x, y) : I(x, y) \geq h\}. \quad (6)$$

4) *Volume Filtering—Volume Operators*: A combination of the above contrast and size connected operators yields the volume reconstruction operator. Volume operators remove connected components from the image whose volume is below a certain threshold. They are defined as

$$\beta_n^-(I)(x, y) = \sup \{h : (x, y) \in \beta_n^-(T_h(I))\} \quad (7)$$

$$\beta_n^+(I)(x, y) = \sup \{h : (x, y) \in \beta_n^+(T_h(I))\} \quad (8)$$

where, in the binary case, if $T_h(I) = X = \bigcup_i X_i$ we define $\beta_n^-(X) = \bigcup \{X_j : \text{Area}(X_j) \cdot h \geq n\}$ with X_j being connected components. Volume operators preserve the formal properties of openings and closings and can be used as a mean of simplification filtering that balances contrast and size criteria.

B. Image Decomposition

In many problems of image analysis, the observed image I may contain texture and/or noise. Texture can be defined as deterministically or stochastically repeated patterns of small scale details, whereas noise is also a pattern of small scale details, but of random, uncorrelated values. The simplification step extracts the most meaningful information from I , resulting to a cartoon-like image U with bounded variation that retains all underlying structure, formed by homogeneous regions with sharp boundaries. In other words, the image U retains all contrast and boundary information but loses all small scale patterns details, which may otherwise be very important especially if they represent texture. In many image analysis tasks, there is a tendency to decompose the image into a sum of functions with different basic characteristics, such as contrast, texture and possibly noise. The resulting model is a linear combination of the form: $I = U + V + W$, where I is the observed image, U is a

simplified sketchy approximation of I , V is the texture component and W is noise [25]. In denoising applications, U is computed as a smooth approximation of I [26]–[28], whereas the rest of the components are not kept. In other cases, both U and V are extracted from I via a constraint variational minimization framework [28], [25], [29].

Based on the aforementioned image decomposition model, given the image I and a simplification method to obtain U (as explained in Section II-A), we can theoretically extract the texture component V . Initially, we perform an image denoising scheme that does not affect image structure such as ASF filtering based on reconstruction, described in Section II-A1. The output image is theoretically free from noise and can be considered as the linear combination of the contrast U and texture V components. Therefore, the texture component can be obtained as the image residual: $V = I - U$. Several nonlinear edge-preserving image smoothing schemes can create cartoon approximations of an image such as anisotropic diffusion and image selective smoothing [27], [30]. To obtain the cartoon component U , we apply the leveling operator of (3) on the initial image, $U = \Lambda(M|I)$, motivated by its interesting properties and the fact that levelings are already employed in the image simplification stage. We construct multiscale cartoons U_i by using multiscale levelings [20] based on a sequence of multiscale markers M_i , obtained from sampling a Gaussian scale-space. The corresponding residuals $V_i = I - U_i$ constitute a hierarchy of multiscale texture components. Additionally, anisotropic diffusion [27] can be considered as an alternative marker selection method. At each sequence step the leveling marker is obtained by a version of the image with blurred regions but adequately preserved boundaries, caused by the constrained diffusion process.

Other alternative schemes for image decomposition can be found in [31] and [32]. It should be noted that the scope of this paper is not to find the optimum $U + V$ decomposition as done in [25] and [29], but some efficient decomposition scheme that produces the texture component V and couples with segmentation methods. An example of this image decomposition scheme can be seen in Fig. 2.

C. Feature Extraction

The feature extraction stage deals with the extraction of special image features that facilitate the final segmentation step and requires a more severe, but detailed, processing of the image. As *features*, we denote regions of interest, gradients, texture measurements, described as follows.

1) *Gradient Features*: High values of the image's gradient are indicative of abrupt intensity changes and specify possible object/region contours. Additionally, the topographic relief, emerging from the gradient magnitude function is used in the flooding process, which, in turn, leads to the final segmentation map. Many different types of gradients have been extensively used in the edge detection framework. Among them, we choose the morphological gradient for its robust behavior, low complexity, and better segmentation results compared to other edge strength operators, such as $||\nabla G_\sigma * I||$. The morphological gradient is computed as the magnitude of the dilated version of the

GEOMETRIC FILTERING ALGORITHM

Geometric filtering (such as Area or Volume openings and closings) is achieved based on filtering the threshold sets $T_h(I)$ of the grey level image $I(x, y)$ and reconstructing the new image from threshold superposition.

A. Analysis Phase

- 1) Obtain all threshold sets $T_h(I)$ for $h = 1, 2, \dots, 255$ (assuming 8-bit/pixel images), as in (6).
- 2) Filter all threshold sets by operating on their connected components: For $h = 1, 2, \dots, 255$
 - Find all disjoint connected component sets X_i of $T_h(I)$, where $T_h(I) = \bigcup_i X_i$ (using 4-connectivity for foreground and 8-connectivity for background).
 - Filter each threshold set $T_h(I)$:
 - *area opening*: check all connected components and reject those X_i that do not satisfy the inequality: $\text{Area}(X_i) \geq n$.
 - *volume opening*: check all connected components and reject those X_i that do not satisfy the inequality: $\text{Area}(X_i) \cdot h \geq n$.
 - Keep all the remaining connected components X_j of $T_h(I)$, thus obtaining area/volume opening at level h .
 - *area opening*:
 $\alpha_n^-(T_h(I)) = \bigcup_j X_j : \text{Area}(X_j) \geq n$.
 - *volume opening*:
 $\beta_n^-(T_h(I)) = \bigcup_j X_j : \text{Area}(X_j) \cdot h \geq n$.

B. Reconstruction Phase

- 1) Reconstruct grey level filtered image by applying superposition of threshold sets:
 - *area opening*:
 $\alpha_n^-(I)(x, y) = \max\{h : (x, y) \in \alpha_n^-(T_h(I))\}$.
 - *volume opening*:
 $\beta_n^-(I)(x, y) = \max\{h : (x, y) \in \beta_n^-(T_h(I))\}$

Fig. 1. Geometric filtering algorithms: Area and volume opening.

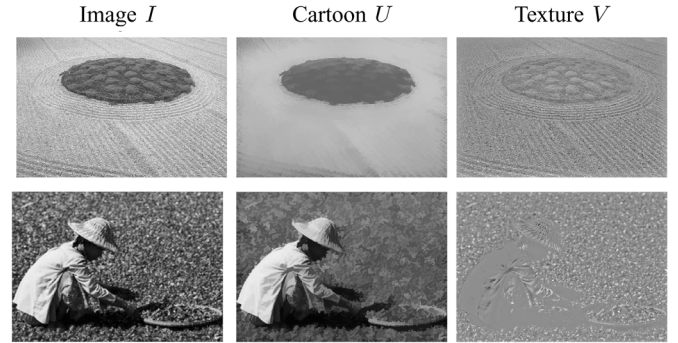


Fig. 2. Image decomposition $I = U + V$ using levelings.

image (after simplification) minus the eroded version of it, normalized by the diameter of the elementary structuring element used in the morphological operations (that is, a 9-pixel square or a 5-pixel diamond, having a diameter equal to 2). This is expressed mathematically as $\mathcal{M}_\nabla(I) = [(I \oplus B) - (I \ominus B)]/2$.

2) *Texture Features*: In Section II-B, we delineated a way of acquiring texture information from the observed image, under an image decomposition scheme. Although all appropriate information is theoretically contained in texture component V , it is in a rough and unprocessed form. Therefore, we need a texture analysis and modeling scheme capable of quantifying texture characteristics.

A way to model and analyze the existing texture patterns of an image is by assuming they are narrowband 2-D AM-FM signals [33], [34] of the form $\alpha(x, y) \cos[\phi(x, y)]$. For wideband signals $f(x, y)$, we assume that they are a sum of 2-D nonstationary sines $f(x, y) = \sum_k \alpha_k(x, y) \cos[\phi_k(x, y)]$, with a spatially varying amplitude $\alpha_k(x, y)$ and a spatially varying instantaneous frequency vector $\vec{\omega}_k(x, y) = (\omega_{k1}(x, y), \omega_{k2}(x, y)) = \nabla \phi_k(x, y)$. In particular, the amplitude is used to model local image contrast and the frequency vector contains rich information about the locally emergent spatial frequencies. An efficient way to estimate the 2-D amplitude and frequency signals was developed in [33] based on the energy operator $\Psi(f) \triangleq \|\nabla f\|^2 - f \nabla^2 f$, which is a multidimensional extension of the 1-D Teager energy operator. Applying Ψ to a 2-D AM—FM signal yields $\Psi[\alpha_k \cos(\phi_k)] \approx \alpha_k^2 \|\vec{\omega}_k\|^2$, which equals the product of the instantaneous amplitude and frequency magnitude squared and may be called the *texture modulation energy*. It should be noted here that Ψ is not directly applied on the wideband signal V , but on narrowband versions of it, filtered through a 2-D Gabor filterbank [35]. The 2-D Gabor filters are characterized by impulse response of the form $h_k(x, y) = e^{-\alpha^2 x^2 - \beta^2 y^2} \cos(\Omega_{k1}x + \Omega_{k2}y)$, where $\alpha/2\pi, \beta/2\pi$ are the rms bandwidths in each dimension and $(\Omega_{k1}, \Omega_{k2})$ is the k -th filter's central frequency pair. The filters are uniformly arranged in the spatial frequency domain, in a polar wavelet-like tessellation, with equal and directional symmetric bandwidths and cover densely the frequency domain. Representations indicative of the dominant texture components are obtained by an energy tracking mechanism in the multidimensional feature space consisting of the filter responses. The filtered texture components are subjected to energy measurements via the 2-D Energy Operator Ψ . The energies are then averaged by a local averaging filter h_a and are subjected to pixelwise comparisons. The *maximum average teager energy*, given by $\Psi_{\text{mat}}[f(x, y)] = \max_k h_a * \Psi[f * h_k](x, y)$ (where $*$ denotes convolution) is kept, as a means of tracking the most active texture component. The derived Ψ_{mat} is a slowly varying indication of texture modulation energy, which can be classified among different energy levels. It provides both local and global texture information and tracks the most dominant texture components along multiple modulation bands [36]. The derived image textural feature is capable of quantifying important characteristics like geometrical complexity, rate of change in local contrast variations and texture scale.

3) *Markers*: Markers are predefined image locations that serve as starting points of the region-growing procedure. They grow in time according to a set of specified criteria until the image plane is totally covered by them. In the case of watershed-like region growing, the marker set is a binary set, where each connected component corresponds to a region of interest. Since markers only serve as starting points of the growing procedure, we are only interested in their location and not their shape. However, their location must be exact corresponding to regions and objects of interest and not point to insignificant structures or texture edges.

In watershed-like region growing, the simplest markers that can be employed are the regional minima of the gradient image. However, it is often the case that the minima are extremely nu-

merous, leading to image oversegmentation, which is considered as problematic. Hence, in most practical cases, the watershed will take as sources of the flooding a smaller set of markers, which have been identified by a preliminary analysis step as inside particles of the regions or objects that need to be extracted via segmentation. The advantage of the aforementioned marker adaptation is robustness, since the result is independent of the shape or the placement of the markers in the zones of interest.

In the literature, there are several marker extraction methodologies, depending on the application. It is common practice that markers are chosen as region where some homogeneity criterion is constant or a key characteristic is of certain strength. Markers can be chosen via a linear methodology to correspond to image key points, such as blobs, ridges, and corners, that remain scale invariant [37], [38]. Alternatively, they can be extracted in a nonlinear way as flat/quasi-flat zones of the morphological gradient function [18] and morphological scale-invariant features.

In our research work, we follow the direction of nonlinear morphological marker extraction methodologies, motivated by their edge, contour and geometric feature preservation properties, their rigorous formulation, computational simplicity, fast implementation and performance that is better or similar to other schemes. Depending on the nature of image under study, we emphasize on contrast, volume and texture-based markers, i.e., image areas where the homogeneity criterion is contrast, volume (area and contrast) and texture, respectively. In all three cases, we extract markers via a reconstruction procedure as valleys or peaks of a image transform that resembles one of the aforementioned characteristics. In all cases, the scale is incorporated in the structuring element or reconstruction controlling parameter. Specifically, we distinguish the following cases.

- *Contrast, Area, or Volume-Based Markers*. Markers are estimated as valleys (or peaks) of certain strength of a generalized bottom (top) hat transform defined as: $H_B(I) = \varphi(I) - I$, where $\varphi(I)$ is a generalized closing and I is an intensity image (initial or simplified). Similarly, the top hat transform is $H_T(I) = I - \gamma(I)$ where $\gamma(I)$ is a generalized opening. Depending on what kind of closing/opening transform we choose, we obtain: a) *contrast* markers if the generalized closing is based on reconstruction, i.e., $\varphi(I) = \rho^+(I + h | I)$ that is where the parameter h controls the contrast (valley depth); b) *area* markers if $\varphi(I)$ is area closing; c) *volume* markers if $\varphi(I)$ is volume closing, in which case contrast and area criteria are exploited.
- *Texture-Based Markers*. Again, markers are estimated as peaks of an image transform that relies on texture characteristics. Ways to compute this transform were illustrated in Section II-B, where we obtained a texture component via an image decomposition model, whereas, in Section II-C2, texture presence was exploited via a multiband energy operator. Therefore, peaks (valleys) either of the texture component V or its dominant modulation energy are extracted as highly (poorly) textured regions. The peak (valley) extraction is based on a reconstruction procedure as discussed earlier.

III. GENERALIZED FLOODING PDE SEGMENTATION

In mathematical morphology, the watershed transform is viewed as the flooding process of the topographic surface of

the image, where the intensity defines the altitude of the relief. Flooding is achieved by immersing the surface into water, letting the water invade through pierced minima. The underlying idea is the following: a) a gradient image of the scene is constructed; b) for each object of interest or homogeneous region, a marker is detected, either in an automatic or interactive manner; c) the watershed lines associated to the markers are constructed. The flooding process of image gradient surface starts with waves emanating from the set of markers, and at points where the waves meet each other, the segmentation boundaries are formed.

Based on the criteria governing the flooding process, different types of segmentation can occur with varying characteristics in their results. By the term *flooding criterion*, we refer to the characteristic that all lakes (associated with the flooding sources) share with respect to water, such as altitude/height (contrast criteria), area (size criteria), or volume (contrast and area criteria) [39]. The most common type of flooding that forms the basis of all traditional morphological segmentation schemes is obtained when the water altitude variation is at the same level for all lakes, known as uniform height flooding. In the next paragraphs we investigate the classical case of *uniform height watershed flooding* and we introduce the case of *uniform volume watershed flooding* modeled via PDEs, where contrast and size criteria are balanced with respect to the topographic relief. In both cases, we derive the corresponding PDE formulations of the floodings viewed as curve evolution processes. Whereas uniform height watershed flooding has already been studied in the continuous domain and modeled via PDEs [4]–[6], we consider uniform volume watershed flooding PDE formulation to be an innovation introduced by our research work. For scientific consistency and smooth transition purposes, we next describe both cases of flooding, emphasizing on PDE formulation and implementation of uniform volume flooding.

A. Uniform Height Flooding

We first consider the case of flooding the 1-D function f shown in Fig. 3(a). We pierce this function at one of its regional minima, as shown in Fig. 3(a), and then we immerse it in water with constant vertical speed. In this case the water altitude at each time instance is uniform. Let ΔH be the height difference from one time instance to another. When the flooding of the function f is done with uniform height speed, we have $(\Delta H)/(\Delta t) = \text{const} = c$. Given that $\tan(\theta) = (\Delta H)/(\Delta L) = |f'|$, we have

$$\frac{\Delta H}{\Delta t} = |f'| \frac{\Delta L}{\Delta t} \Rightarrow v = \frac{\Delta L}{\Delta t} = \frac{c}{|f'|}. \quad (9)$$

By the term v , we refer to the horizontal velocity, by which the level sets of the function f propagate in time.

Let us now consider the case of a 2-D function. Shown in Fig. 3(b) is the planar projection of a lake of a 2-D function, flooded under the constraint of uniform height. The boundary of the lake at time instance t is the set $\Gamma(t)$ of points of the closed planar curve represented by its position vector $\vec{C}(t)$. This curve models the propagation of a wave emanating from the lake. Based on (9) and considering that in the 2-D case i) ΔL

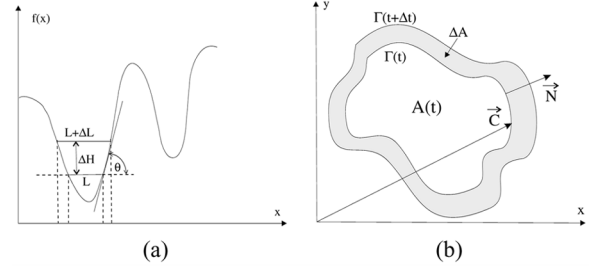


Fig. 3. (a) Lakes of 1-D function; $L(t)$ is the length of level sets. (b) Planar projection of a lake of 2-D function; $A(t)$ is the area of level sets.

becomes the displacement normal to the level curves of the function and ii) $|f'|$ becomes $\|\nabla f\|$, we model the propagation of the curve \vec{C} with the following PDE:

$$\frac{\partial \vec{C}}{\partial t} = \frac{c}{\|\nabla f\|} \cdot \vec{N} \quad (10)$$

where \vec{N} is the unit outward normal vector to the curve.

Following the level set formulation proposed by Osher and Sethian [40], we embed this evolving planar curve as the zero-level curve of an evolving space-time function $\Phi(x, y, t)$; i.e., $\Gamma(t) = \{(x, y) : \Phi(x, y, t) = 0\}$. Then the PDE that governs the evolution of the obtained level function is

$$\frac{\partial \Phi}{\partial t} = v(x, y) \|\nabla \Phi\| \quad (11)$$

where $v(x, y)$ is the space-dependent speed function given by $v(x, y) = c/\|\nabla f(x, y)\|$.

B. Uniform Volume Flooding

Considering the case of uniform volume speed flooding, the main difference, compared to traditional watershed segmentation, is that during the flooding process the water height is not at same level for all lakes. However, the water volume change rate is constant, thus retaining the balance between area and contrast characteristics. Following the same procedure as above we flood the 1-D function f of Fig. 3(a). We then have that $L(\Delta H)/(\Delta t) = \text{const} = c$, which yields the horizontal velocity v

$$\frac{\Delta L}{\Delta t} = \frac{\Delta H}{\Delta t} \frac{1}{|f'|} \Rightarrow v = \frac{\Delta L}{\Delta t} = \frac{c}{L(t) |f'|}. \quad (12)$$

In the 2-D case of Fig. 3(b), where the curve \vec{C} models the wave emanating from a lake flooded under the constraint of uniform volume speed $L(t)$ becomes $\text{Area}(t)$, which is the area enclosed by the propagating wave at time t . The curve evolution PDE and the equivalent PDE level function are

$$\frac{\partial \vec{C}}{\partial t} = \frac{c}{\text{Area}(t) \|\nabla f\|} \cdot \vec{N} \quad (13)$$

$$\frac{\partial \Phi}{\partial t} = v(x, y, t) \|\nabla \Phi\| \quad (14)$$

where $v(x, y, t) = c/(\text{Area}(t) \|\nabla f(x, y)\|)$ is the time and space-dependent speed function. The product $\text{Area}(t) \|\nabla f\|$ is a measure of volume; thus, the speed of the evolving curve is inversely proportional to the volume of the sources of the flood

and the flood is slowed down by the factor $1/v$. This means that during the flooding process lakes with large volume are filled up by water slowly whereas lakes with small volume are filled up quickly. The evolution PDE (13) can be interpreted as a dilation of the front with a varying radius structuring element. The radius $1/(\text{Area}(t)\|\nabla f(x,y)\|)$ of the structuring element depends on space (x,y) and time t .

C. Implementation Details and Experimental Results

The PDEs (10) and (13) are time-dependent and since in both cases the evolving front's speed is one-directional (expanding fronts), we can consider the stationary formulation of the embedding level function evolution PDEs (11) and (14) with positive speed v , which is known as *eikonal* PDE. The problem then is to solve the equation $\|\nabla T(x,y)\| = 1/v$ where $T(x,y) = \inf\{t : \Phi(x,y,t) = 0\}$ is the *minimum time of arrival*, at which the zero-level curve of the level function $\Phi(x,y,t)$ crosses (x,y) . In the two cases of flooding discussed earlier, the resulting "stationary" eikonal-type PDEs are

$$\begin{aligned} \text{Height Flooding : } \|\nabla T(x,y)\| &= \|\nabla f\|/c \\ \text{Volume Flooding : } \|\nabla T(x,y)\| &= \text{Area}(t)\|\nabla f\|/c. \end{aligned} \quad (15)$$

A fast algorithm used to solve such PDEs is the FMM [41], which provides an extremely fast way for solving eikonal equations. A similar algorithm was developed in the context of time-optimal control [42].

The FMM methodology was adapted to accommodate our scenario of multiple competing propagating curves that correspond to boundaries of expanding markers. The propagation of all curves is done simultaneously, under the constraint that a region covered by a front cannot be infiltrated by a different front. The evolving fronts compete in the sense that a pixel is assigned to the front that claims (reaches) it first. At places where two or more fronts meet, a dam is erected to specify the segmentation line. The multiple fronts propagate in time until convergence is met when all pixels in the image domain have been assigned to a front. Whereas in the case of uniform height flooding FMM is easy to implement as done in [6], in the case of uniform volume flooding solving the PDE $\|\nabla T(x,y)\| = (1/c)\text{Area}(t)\|\nabla f\|$ has the peculiarity of the pseudo time varying term $\text{Area}(t)$. This is why the fast marching algorithm has to be adapted so as to take under consideration the multiple curve scenario as well as the time-dependent area variations of the fronts. Every time we "march" a front one pixel forward, its area is incremented; thus, the corresponding area term has to be updated.

Experimental results of height and volume flooding segmentation are illustrated in Fig. 5. The images shown in Fig. 5(a) and (e) are synthetically produced by taking the distance transform of the corresponding binary images and adding an arbitrary constant to each of their connected components, thus producing graylevel images. For illustration purposes, on each synthetic image we have superimposed a flooding source for each object, shown in white color. Bright objects have high contrast/altitude compared to darker objects. In Fig. 5(b), the contour lines of each object are presented, with blue color corresponding to lower altitude and red corresponding to higher altitude. The cases of uniform height and volume flooding are ex-

FAST MARCHING ALGORITHM ADAPTATION

Initially the binary set of markers consists of K connected components, each assigned a label in $\{1, \dots, K\}$. Each component's boundary corresponds to an evolving curve. Pixel locations are indexed on a grid (i,j) .

A. Initialization Phase

- 1) Create the **Alive** set that consists of all pixels for which their arrival time equals zero (that is the pixels enclosed by the propagating curves).
- 2) Create the **NarrowBand** set that consists of the points 1 pixel far from the boundary (propagating curves), setting their arrival time equal to $T_{ij} = 1/v$
- 3) Create the **FarAway** set that consists of all remaining pixels, setting their arrival time $T_{ij} = \infty$

B. Marching Phase

- 1) Find (i_{\min}, j_{\min}) point in NarrowBand with minimum time of arrival T .
- 2) Move (i_{\min}, j_{\min}) from NarrowBand to Alive set.
- 3) Set the label of pixel (i_{\min}, j_{\min}) equal to the label of its alive neighbor that pushed it in Narrowband set (either in step A.2 or in step B.5).
- 4) Update the area of the front the (i_{\min}, j_{\min}) has been assigned to.
- 5) Find neighbors $(i_{\min} - 1, j_{\min}), (i_{\min} + 1, j_{\min}), (i_{\min}, j_{\min} - 1), (i_{\min}, j_{\min} + 1)$ that do not belong to Alive.
 - If these neighbors belong to FarAway set move them to NarrowBand.
- 6) Recompute T for all neighbors by solving equation: $[\max(D_{ij}^{-x}T, 0)^2 + \min(D_{ij}^{+x}T, 0)^2 + \max(D_{ij}^{-y}T, 0)^2 + \min(D_{ij}^{+y}T, 0)^2]^{1/2} = 1/v$
- 7) Return to step 1 of Marching Phase.

Fig. 4. Adaptation of FMM algorithm.

amined. Considering the synthetic image of Fig. 5(a), and given the four markers (flooding sources) scenario, one placed on each component, we expect that after the segmentation the image will be partitioned in four regions. Since the background is an extra region, we expect that one object will be eventually merged with the background region. We investigate the lost-object scenario for each type of flooding. In the case of height flooding, the object of lowest contrast is totally lost. Additionally, the boundaries of the remaining objects are not accurately extracted. In the case of volume flooding, the undetected object is the one of lowest volume (area and contrast), but the boundaries of the remaining objects are correctly traced. Similar conclusions are drawn by observing the results in Fig. 5(g) and (h). Whereas in Fig. 5(a), the objects of lower contrast and lower volume are the same, in Fig. 5(e), that is not the case. The presented experiments exploit the basic property of volume flooding i.e., retaining the balance between area and contrast, and illustrates its advantage against height flooding in such cases.

IV. COUPLING GEOMETRY AND TEXTURE IN SEGMENTATION

Up to this point, we have focused on watershed-like segmentation implemented via PDEs. In the cases investigated above, the characteristic under consideration that controlled the growing process was image intensity viewed either as seeds' contrast, size or a combination of these two (volume). However, most of real word images do not only consist of contrasted areas but they are also composed of textured parts. Trying to segment such images based on purely contrast information, captured by gradient operators, leads to undesirable results, since intensity

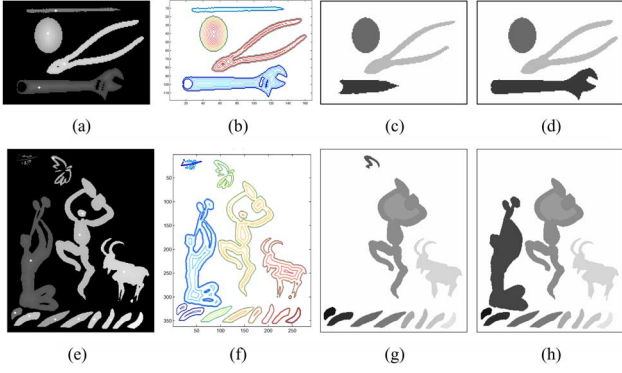


Fig. 5. Height and volume flooding segmentation results. (a), (e): Synthetic images; (b), (f): contours corresponding to different altitudes (gray values); (c), (g): height flooding segmentation regions; (d), (h): volume flooding segmentation regions.

variations in amplitude and phase, owing to textured parts, introduce false edges that affect the final segmentation result. Based on evidence from psychophysics, according to which humans combine multiple cues in order to detect boundaries from images [43], and in view of the fact that contrast and texture are two distinct cues bearing different type of information, we try to extract and use them separately so as to improve and balance the segmentation results.

Based on the PDE of (13), which achieves either volume ($\text{Area}(t) \neq 1$) or height flooding ($\text{Area}(t) = 1$) by exploiting intensity contrast information, we attempt adding to it an extra term that will control the image textural information. Ideally we want the extra term to enable the growing seeds surpass false edges introduced by texture structures in the image, by speeding up the evolution at such places. In other words, we want an extra texture controlled term that will be able to quantify properly the available image texture information. Now recalling the derived texture features of Section II-C2, we realize that Ψ_{mat} operator provides both local and global texture information, tracks the most dominant texture components along multiple modulation bands and is capable of quantifying important characteristics like geometrical complexity, rate of change in local contrast variations and texture scale.

As a result, the desired extra term, which bares texture information in such a way that favors region growing within textured parts, is the multiband energy operator of Section II-C2. A way to incorporate contrast and texture as different cues in a segmentation scheme is by modifying the (13) in the following way:

$$\frac{\partial \vec{C}}{\partial t} = \left(\frac{\lambda_1}{\text{Area}(t) \|\nabla I\|} + \lambda_2 \Psi_{\text{mat}}(I) \right) \vec{N} \quad (16)$$

where λ_1 and λ_2 are parameters that control the contribution of each cue. Thus, the seed evolution speed depends on two eikonal terms, linked with some optimality criterion. The first term drives the curve (seed's boundary) with speed that maximizes the *flooding* of the image toward its watershed. The second term can be shown to correspond to a flow that maximizes the average texture energy: $\max \int_{R(C)} \Psi(I) \Rightarrow \partial \vec{C} / \partial t = \Psi(I) \vec{N}$. This term pushes the curve toward regions with large average texture energy.

Particularly, the segmentation scheme of (16) can be written in the following generalized form:

$$\frac{\partial \vec{C}}{\partial t} = \left(\frac{\lambda_1}{\max(\epsilon, \text{Area}(t) \|\nabla I\|)} + \lambda_2 \Psi_{\text{mat}}(I) \right) \vec{N} \quad (17)$$

where $0 \leq \epsilon \leq 1$ is used to handle instabilities caused by gradient's zero values.

A. Coupled $U + V$ PDE

The previously derived PDE of (16) consists of two components: one that controls contrast information using gradient magnitude operator and another that controls texture information by texture modulation energy operator. Although the gradient operator quantifies intensity changes and the energy modulation operator quantifies AM-FM variations corresponding to texture, each one's privilege is eliminated by the application on the same input image I . However, if these two operators could be applied on separate image transformations emphasizing on different type of information, the scheme of (16) would be more efficient and powerful.

The recently proposed image decomposition model [25], [29] constitutes a solution to the aforementioned problem, thus providing an effective way of linearly distinguishing contrast and texture from a single image, in the form $I = U + V$. Specifically, the U component, known as *cartoon*, serves very well as a contrast descriptor since it consists of relatively flat plateaus that correspond to object regions, surrounded by abrupt edges that correspond to object boundaries. The V component, which is in fact the *texture oscillation* contains texture plus noise information and serves as texture descriptor. Combining the $U + V$ image decomposition philosophy with the PDE of (16) we conclude to a coupled segmentation PDE where contrast variations are taken into account from the U part and texture oscillations are approached through modulation analysis on the V component

$$\frac{\partial \vec{C}}{\partial t} = \left(\frac{\lambda_1}{\text{Area}(t) \|\nabla U\|} + \lambda_2 \Psi_{\text{mat}}(V) \right) \vec{N}. \quad (18)$$

The latter is a curve evolution with speed inversely proportional to the intensity contrast (or volume) of the cartoon component and proportional to the Ψ_{mat} energy of the textured component. We obtain the cartoon component U by applying the *leveling* operator [20] on the initial image, and the texture component as the residual $V = I - U$. Treating and processing the two components separately and combining them for the image partitioning part results in a powerful joint segmentation scheme of (18).

It should be noted that geometric curve evolution of the form $\partial \vec{C} / \partial t = g(c - \mu \kappa) \vec{N}$ has been proposed by Caselles *et al.* [44] and Malladi *et al.* [45]. Our proposed scheme has three differences from the aforementioned geometric evolution: i) the first term achieves watershed type flooding, ii) the second term is a new contribution that acts on the texture component of the image, and has never been used before in segmentation schemes, and iii) the curvature component κ is not present in our scheme since it was experimentally determined that it does not provide any significant improvement to the overall segmentation.

For the implementation of the proposed curve evolution scheme, we follow the level set formulation [40], and we conclude to the level function PDE

$$\frac{\partial \Phi}{\partial t} = \left(\frac{\lambda_1}{\text{Area}(t) \|\nabla U\|} + \lambda_2 \Psi_{\text{mat}}(V) \right) \|\nabla \Phi\|. \quad (19)$$

The curves that propagate according to (19) are multiple, initialized as the contours of a set of markers, thus indicating significant image regions. The marker extraction is done according to the methodologies described in Section II-C. Specifically, depending on the type of image to be segmented we choose our markers to be contrast oriented, texture oriented, a combination of the above, or manually placed at areas of interest. It should be clarified here that the proposed segmentation method including marker extraction is fully automated. However, marker extraction can be operated with human intervention, according to the needs of each application. The implementation of (19) has been done with established techniques from level sets methods. Specifically, the PDE (19) is of pure eikonal-type and its implementation is based on the FMM [41], [42], in the way it was described in Section III-C, and ensures computational speed. Elimination of division-by-zero scenario as in PDE (17) can be also applied in PDEs (18) and (19) in order to handle instabilities caused by cartoon gradient zero values. Automatic and human intervention-based segmentation examples are illustrated in Figs. 6 and 7. In Fig. 6, we demonstrate a set of the extracted features and segmentation results on a natural image. The manually placed markers overimposed on the image are shown in Fig. 6(a). In Fig. 6(b) and (c), we illustrate the U and V image components obtained after image decomposition, and in Fig. 6(d) and (e), we present the (inversed) modulation energies $\Psi_{\text{mat}}(I)$ of the image I and of the texture component $\Psi_{\text{mat}}(V)$. The segmentation result using the PDE of (17) is shown in Fig. 6(f), whereas the results of using PDE (18) is shown in Fig. 6(g). In Fig. 6(h), we illustrate the watershed segmentation result for visual comparison. As it can be seen in Fig. 6(g), which is the case of U and V coupled segmentation scheme, region boundaries are better localized and there is no false contour placement due to texture edges. The incorporation of U and V image components in the PDE segmentation scheme outperforms the results of the other two schemes. Another application of a soilsection image segmentation is illustrated in Fig. 7. Such images demonstrate highly textured and contrasted areas, and have served as a strong motivation for us in our effort to introduce and exploit the coupled segmentation scheme.

B. Adaptation of λ Parameters

So far, we have introduced a PDE-based segmentation scheme that combines contrast, size and texture criteria. Each of these cues' contribution is controlled by a coefficient, namely λ_1 contrast/size evolution controlling parameter and λ_2 texture evolution controlling parameter. For the computation of these parameters we can distinguish the following scenarios: i) λ_1 and λ_2 remain spatially constant and ii) λ_1 and λ_2 remain spatially varying. Obviously, the second case seems to be more efficient due to its ability to adapt depending on local image characteristics, whereas the first case takes under consideration global image criteria.

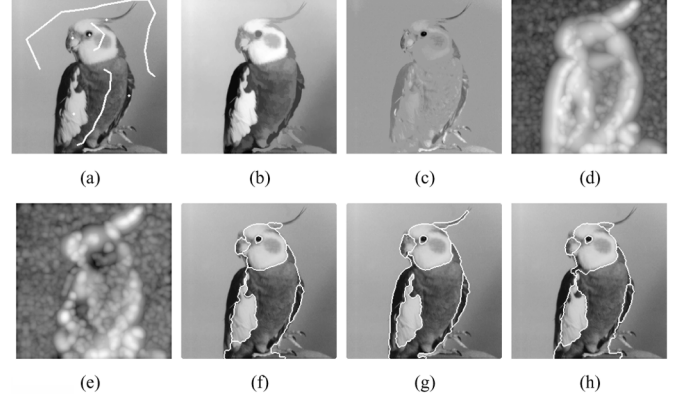


Fig. 6. Image features and segmentation results: (a) original image and markers; (b) cartoon U ; (c) texture V ; (d) texture modulation energy $\Psi_{\text{mat}}(I)$ (log plot); (e) texture modulation energy $\Psi_{\text{mat}}(V)$ (log plot); (f) coupled segmentation on I ; (g) coupled segmentation on $U + V$; (h) watershed flooding segmentation on I .

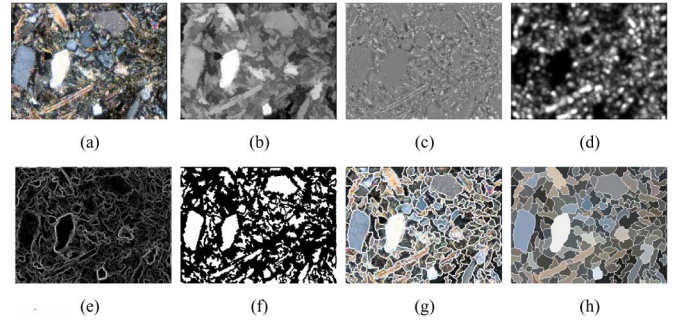


Fig. 7. Image features and segmentation results: (a) soilsection image I ; (b) cartoon U ; (c) texture V ; (d) texture modulation energy $\Psi_{\text{mat}}(V)$; (e) $\tilde{\mathcal{M}}(I)$; (f) markers; (g) segmentation result; (h) segmentation mosaic.

In the case of spatially constant λ , we assume $\lambda_1 + \lambda_2 = 1$, and we examine a wide range of possible λ pairs. For each pair, we compute the corresponding segmentation map and for each map we compute a quality criterion. The pair that maximizes (minimizes) the criterion constitutes the optimal choice for λ parameters. Obviously, this is a computational demanding method, and the only reason of using it is as a reference for comparative studies. In the case of spatially adaptable λ parameters, we take advantage of the fact that the already used $U + V$ image decomposition model gives evidence about the existence of each component (contrast and texture) at every image location. As it can be observed in (18), all the needed information about contrast at each image pixel is encapsulated by $1/|\nabla U|$ component. Similarly, all information about texture contribution is captured by $\Psi_{\text{mat}}(V)$. Hence, we estimate λ_1 (contrast coefficient) and λ_2 (texture coefficient) as the mean square error between the observed image I and the texture V or contrast U component, respectively, weighted locally by a small Gaussian window $G_\sigma(x, y)$ of scale σ , i.e.,

$$\begin{aligned} \lambda_1(x, y) &= [G_\sigma * (I - V)^2](x, y) \\ \lambda_2(x, y) &= [G_\sigma * (I - U)^2](x, y). \end{aligned} \quad (20)$$

We can either use this estimated λ -space functions directly or normalize their sum to 1. Alternatively, the coefficients can

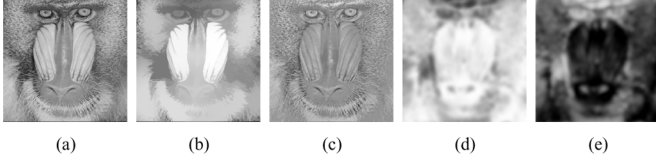


Fig. 8. Spatially varying λ s: (a) mandrill image; (b) cartoon U ; (c) texture V ; (d) spatial visualization of parameter λ_1 ; (e) spatial visualization of parameter λ_2 .

be estimated as: $\lambda_1(x, y) = \exp(-[G_\sigma * (I - U)^2](x, y))$ and $\lambda_2(x, y) = \exp(-[G_\sigma * (I - V)^2](x, y))$. The former selection of λ parameters has experimentally been found to yield slightly better results.

Experimental results have shown that the use of spatially varying λ functions improves the overall segmentation result. In Fig. 8, we present graphically the original image I , its U and V components and the corresponding spatially varying λ functions. The optimum constant pair is $\lambda_1 = 0.3$ and $\lambda_2 = 0.7$ leading to a segmentation result of total region variance over U equal to 0.025, whereas the spatially varying pair of λ_1 and λ_2 of Fig. 8 produces a segmentation result of total region variance over U equal to 0.023.

V. COMPARISONS AND RESULTS

The performance evaluation of a segmentation algorithm is an important subject in the study of segmentation, and it can be done either analytically or empirically [46]. The analytical methods directly examine and assess the segmentation algorithms themselves by analyzing their principles and properties. Empirical methods indirectly judge the segmentation algorithms by applying them to test images and measuring the quality of segmentation results, using i) a *goodness* criterion on the segmented image without *a priori* knowledge of the desired segmentation result and ii) the disparity between the actually segmented image and an ideally segmented reference one, known as ground truth (GT). In our case, we are not concerned with matters of algorithmic complexity nor computational requirements, but instead, we focus on empirical evaluation methods.

In order to judge the quality of segmented images produced by the $U + V$ coupled PDE method, we shall use some quality measures to quantify our results and test them against results produced by other methodologies. There is a variety of goodness measures [46] that have been proposed by many researchers for the evaluation of segmentation methodologies. Most of them are established according to human perception and intuition, and each of them can be used in different segmentation scenarios, but again there is no global measure that can be applied in every case. Among them, we have tested several of them and eventually concluded to following goodness measures.

- *Liu–Yang Global Cost Criterion* [47], which is a tradeoff between suppression of heterogeneity and preservation of details

$$C_{LY} = \sqrt{K} \times \sum_{i=1}^K \frac{E_i^2}{\sqrt{|R_i|}} \quad (21)$$

where K is the total number of regions, E_i is the error over region R_i defined as the Euclidean distance between the original image and the mosaic segmented image, and $|R_i|$ is the area of the region (in pixels). The smaller C_{LY} the better the segmentation results.

- *Mumford–Shah Criterion*: Based on the energy minimization functional [26] $E(g, \Gamma) = \mu \int \int_R (g - I)^2 dx dy + \int \int_{R-\Gamma} \|\nabla g\| dx dy + \nu |\Gamma|$, by eliminating the second term and setting $\mu = \nu = 1$ we conclude to the following criterion:

$$C_{MS} = \int \int_R (g - I)^2 dx dy + \text{Len}(\Gamma) \quad (22)$$

which corresponds to minimization of the average region error and region contour length. g is the segmented mosaic image and Γ is the total segmentation boundary. The smaller C_{MS} the better (g, Γ) segments image I .

- *Texture-Contrast Locally Weighted Mumford–Shah Criterion*: Based on the philosophy of Mumford–Shah evaluation criterion, and going a step further, we introduced a relatively simple goodness measure that takes into account both contrast and texture information and can be viewed as a locally weighted Mumford–Shah criterion. According to this criterion, we measure the variance in each region mapped on a cartoon version of the image $\text{var}(U)$, as well as each region's modulation energy variance $\text{var}(\Psi_{\text{mat}}(V))$. The lower those variance values, the better are the segmentation results. The final measure is a weighted variance obtained as follows. Given the segmentation of the image domain into disjoint regions $R_i, i = 1, \dots, K$, we introduce a weighted Euclidean distance metric that compares two discrete images F and G over the segmentation partition $\{R_i\}$

$$D_{\{R_i\}}(F, G) = \sum_{i=1}^K \frac{1}{|R_i|} \sum_{(x,y) \in R_i} [F(x, y) - G(x, y)]^2. \quad (23)$$

Then, if F is either $U, V, \Psi_{\text{mat}}(V)$, or any other contrast, texture transformations of the input image I , by the term *weighted variance* of the image transform $F(x, y)$, we imply the following:

$$\begin{aligned} \text{var}(F) &= \sum_{i=1}^K \frac{1}{|R_i|} \sum_{(x,y) \in R_i} (F(x, y) - \bar{F}_i)^2 \\ &= D_{\{R_i\}}(F, \bar{F}) \end{aligned} \quad (24)$$

where \bar{F}_i is the mean value of $F(x, y)$ in each R_i and $\bar{F}(x, y)$ is the corresponding mosaic image.

- *Segmentation Boundary Length*: The length $\text{Len}(\Gamma)$ of the segmentation boundaries is calculated as the sum of the perimeters of the regions obtained by the segmentation method.

The choice of segmentation methods, to which the proposed scheme will be tested against, is neither easy, nor obvious. To begin with, the selected methods should produce similar results with the proposed one, that is they should produce closed boundaries and disjoint, plane-filling regions. Ideally,

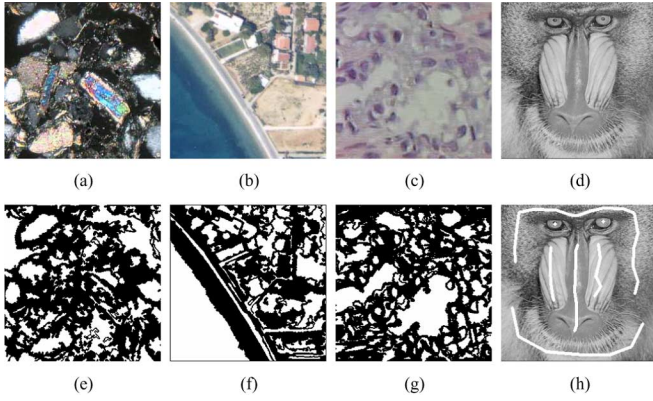


Fig. 9. Reference images and marker sets for comparisons of different watershed-like segmentation methods: (a) soilsection image; (b) aerial image; (c) biomedical image from prostate tissue; (d) mandrill image; (e) marker set; (f) marker set; (g) marker set; (h) marker set.

the number of regions in the two schemes should be the same. We concluded to the selection of the following methodologies: watershed segmentation based on height and volume flooding as well as the multicue scheme without image decomposition of (17). In Fig. 10, we provide a set of different segmentation results obtained by applying the aforementioned methods on four different reference images shown in Fig. 9(a)–(d). The selected images are: 1) a soilsection image consisting of highly contrasted and textured areas, 2) an aerial photo, 3) a biomedical image of prostate tissue, and 4) an animal image with differently textured areas. The different segmentation methodologies are tested using the same set of automatically extracted markers via contrast or volume criteria for images Fig. 9(a)–(c), and a set of manually placed markers for image shown in Fig. 9(d). The corresponding marker sets are illustrated in Fig. 9(e)–(h). Fig. 10(a) illustrates the results of the multicue segmentation scheme without decomposition of (17) applied on each image. Fig. 10(b) demonstrates segmentation results of the proposed multicue scheme with decomposition of (18), whereas Fig. 10(c) and (d) demonstrates segmentation results of height and volume watershed flooding. Apart from visual comparisons, we provide Table I, where the evaluation criteria of (21), (22), (24), and $\text{Len}(\Gamma)$ are computed for each case. As it can be observed in Table I, the proposed multicue coupled scheme with $U + V$ image decomposition performs better than the other methodologies, with respect to computed values of the evaluation criteria. In the case of mandrill image of Fig. 9(d) and the application of Liu–Yang global cost criterion, the performance of watershed volume flooding is better than the one of coupled $U + V$ scheme. This is justified since Liu–Yang global cost criterion is a tradeoff between suppression of heterogeneity and preservation of details, and, hence, it matches better the balanced contrast-area results of volume watershed flooding applied on highly textured images such as the mandrill image. Cases like the above served as rationale for the introduction of the balanced criterion of (24), which incorporates contrast and texture in the evaluation of the segmentation results.

Concluding, we should stress the fact that based on various quantitative results with emphasis on the improved criterion of (24), the proposed multicue coupled scheme with image

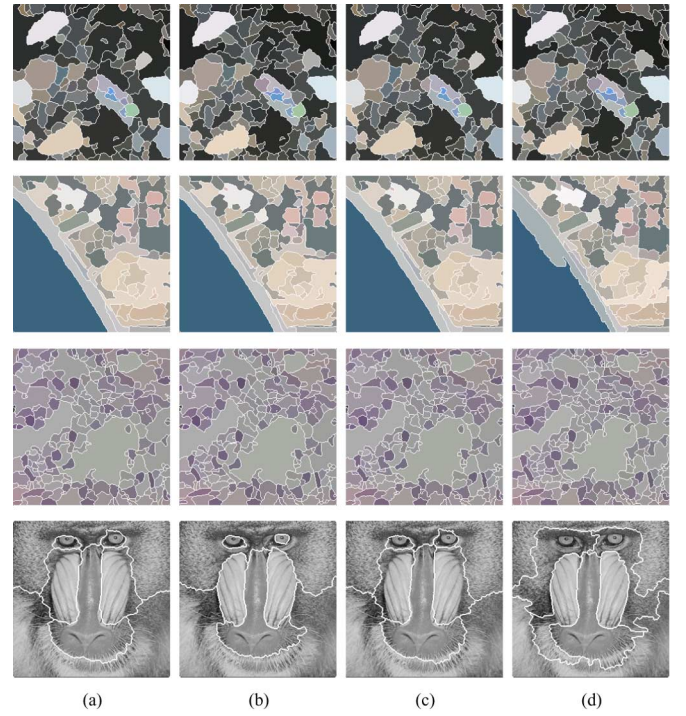


Fig. 10. Comparisons of different types of watershed-like segmentation results: (a) multicue segmentation results without decomposition; (b) multicue segmentation results with decomposition; (c) height watershed flooding segmentation results; (d) volume watershed flooding segmentation results.

TABLE I
EVALUATION COMPARISONS

	Quality Measures	Segmentation Method			
		Coupled Type		Watershed Flooding	
		I	U+V	Height	Volume
soil	C_{LY}	0.417	0.298	0.413	0.314
	C_{MS}	0.178	0.174	0.180	0.184
	$\text{var}(U)$	0.921	0.823	0.893	1.108
	$\text{var}(\Psi_{\text{mat}}(V))$	0.280	0.259	0.281	0.254
	$\text{len}(\Gamma)$	4855	4987	4982	5742
aerial	C_{LY}	0.335	0.322	0.378	0.324
	C_{MS}	0.131	0.132	0.134	0.138
	$\text{var}(U)$	0.335	0.281	0.337	0.383
	$\text{var}(\Psi_{\text{mat}}(V))$	0.473	0.468	0.479	0.555
	$\text{len}(\Gamma)$	3934	4206	4054	4442
biomed	C_{LY}	0.724	0.461	0.718	0.509
	C_{MS}	0.138	0.136	0.140	0.151
	$\text{var}(U)$	0.327	0.294	0.314	0.365
	$\text{var}(\Psi_{\text{mat}}(V))$	0.138	0.135	0.140	0.139
	$\text{len}(\Gamma)$	6529	6630	6728	7593
mandrill	C_{LY}	0.423	0.329	0.420	0.328
	C_{MS}	0.131	0.126	0.132	0.134
	$\text{var}(U)$	0.046	0.024	0.046	0.034
	$\text{var}(\Psi_{\text{mat}}(V))$	0.272	0.232	0.271	0.285
	$\text{len}(\Gamma)$	1167	1210	1201	1960

decomposition outperforms the other segmentation methodologies. It provides better results, in the sense that the resulting partitioning map consists of more uniform regions (low cartoon variance values) with smoother texture (low modulation energy variance), compared to the other methodologies.

Regarding empirical discrepancy methods, we have applied our segmentation methodology on images from the Berkeley segmentation database and used the available reference images for evaluation. All available images originate from real world

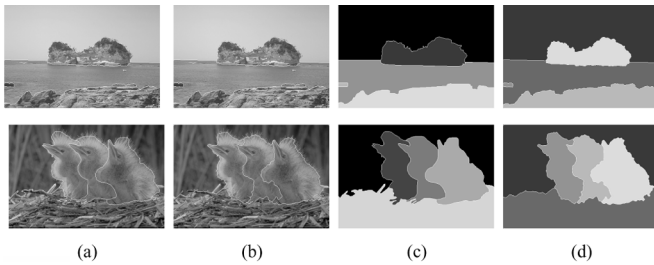


Fig. 11. Comparisons of coupled PDE segmentation results against Berkeley's database GT data. (a) GT segmentation; (b) coupled scheme segmentation result; (c) GT regions; (d) coupled scheme regions.

scenes and the corresponding GT data are a result of manual segmentation by human subjects, which implies that the GT data are subjective to a certain degree of ambiguity. As a measure of distance between the segmented and the GT reference image we have used a region-based consistency measure, the *bidirectional consistency error* (BCE) [48]. Markers are placed manually inside each region of interest. In Fig. 11, we exhibit the segmentation results of our coupled proposed methodology as well as an indicate GT image obtained by human segmentation. The BCE is obtained by comparing our segmentation result to all available GT data (not only the one illustrated here) and taking the average value among them. For the image in the first row of Fig. 11, the BCE is 0.061, interpreted as the mean square error between our segmentation result and the average human segmentation, in terms of region map distance. The above BCE value is very satisfactory, regarding the fact that human segmentations do not bear any accuracy in pixel level. For the image in second row of Fig. 11, the obtained BCE is 0.21 which is quite good, regarding the complicated structure of the image as well as the simplified human segmentation result.

VI. CONCLUDING REMARKS

This paper addressed the problem of image segmentation as a set of subtasks, with emphasis on the image partitioning stage, viewed as a generalized flooding procedure incorporating different information cues. Specifically, the stages of image simplification, decomposition into constituent parts and feature extraction were investigated, and connected operators of different types were proposed as the desirable tools for image presegmentation processing. The classic paradigm of watershed transform was modeled and extended via PDEs using ideas from curve evolution and level sets, in order to satisfy volume flooding criteria. Additionally, using ideas like $U + V$ image decomposition and texture AM-FM modeling, geometrical and textural information were coupled in a novel region growing watershed segmentation scheme. The proposed extended scheme was evaluated in terms of results' quality and was compared to other watershed-like segmentation methods. The obtained qualitative and comparative results verified the fact that the coupling of structural and textural information via decomposition, embedded in a generalized flooding segmentation procedure and modeled via PDEs, leads to high quality results, and improves the classical watershed flooding paradigm. Finally, we note that additional visual cues, such as color, could be added as forces

in our proposed PDE-based coupled segmentation scheme; this is part of our ongoing work in this area.

ACKNOWLEDGMENT

The authors would like to thank G. Evangelopoulos for his valuable help and support in the texture analysis part.

REFERENCES

- [1] F. Meyer and S. Beucher, "Morphological segmentation," *J. Vis. Commun. Image Represent.*, vol. 1, no. 1, pp. 21–45, 1990.
- [2] L. Vincent and P. Soille, "Watershed in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [3] S. Beucher and F. Meyer, E. R. Dougherty, Ed., "The morphological approach to segmentation: The watershed transformation," in *Mathematical Morphology in Image Processing*. New York: Marcel Dekker, 1993.
- [4] L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Process.*, vol. 38, no. 7, pp. 99–112, 1994.
- [5] P. Verbeek and B. Verwer, "Shading from shape, the eikonal equation solved by grey-weighted distance transform," *Pattern Recognit. Lett.*, vol. 11, pp. 618–690, 1990.
- [6] P. Maragos and M. A. Butt, "Curve evolution, differential morphology, and distance transforms applied to multiscale and eikonal problems," *Fundam. Inf.*, vol. 41, pp. 91–129, 2000.
- [7] P. Maragos and M. A. Butt, "Advances in differential morphology: Image segmentation via eikonal PDE, curve evolution and reconstruction via constrained dilation flow," in *Mathematical Morphology and Its Applications to Image and Signal Processing*, H. Heijmans and J. Roerdink, Eds. Norwell, MA: Kluwer, 1998.
- [8] H. Tek and B. B. Kimia, "Curve evolution, wave propagation, and mathematical morphology," in *Proc. 4th Int. Symp. Mathematical Morphology*, 1998, pp. 156–162.
- [9] H. T. Nguyen, M. Worring, and R. Boomgaard, "Watersnakes: Energy-driven watershed segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 3, pp. 330–342, Mar. 2003.
- [10] T. Brox, M. Rousson, R. Deriche, and J. Weickert, "Unsupervised segmentation incorporating colour, texture, and motion," in *Computer Analysis of Images and Patterns*, N. Petkov and M. A. Westenberg, Eds. Groningen, The Netherlands: Springer, 2003, vol. 2756, pp. 353–360.
- [11] J. D. Martin and C. Fowlkes, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.
- [12] L. Shafarenko, M. Petrou, and J. Kittler, "Automatic watershed segmentation of randomly textured color images," *IEEE Trans. Image Process.*, vol. 6, no. 11, pp. 1530–1544, Nov. 1997.
- [13] A. Sofou and P. Maragos, "PDE-based modelling of image segmentation using volumic flooding," presented at the IEEE Int. Conf. Image Processing, 2003.
- [14] A. Sofou, G. Evangelopoulos, and P. Maragos, "Coupled geometric and texture PDE-based segmentation," presented at the IEEE Int. Conf. Image Processing, 2003.
- [15] A. Sofou, G. Evangelopoulos, and P. Maragos, "Soil image segmentation and texture analysis: A computer vision approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 4, pp. 394–398, Oct. 2005.
- [16] J. Serra, *Image Analysis and Mathematical Morphology, Vol II: Theoretical Advances*. New York: Academic, 1988, vol. 2.
- [17] H. Heijmans, *Morphological Image Operators*. New York: Academic, 1994.
- [18] P. Salembier and J. Serra, "Flat zones filtering, connected operators, and filters by reconstruction," *IEEE Trans. Image Process.*, vol. 4, no. 8, pp. 153–160, Aug. 1995.
- [19] F. Meyer, H. Heijmans and J. Roerdink, Eds., "The levelings," in *Proc. 4th Int. Symp. Mathematical Morphology and Its Applications to Image Processing*, Jun. 1998, pp. 199–206.
- [20] F. Meyer and P. Maragos, "Nonlinear scale-space representation with morphological levelings," *J. Vis. Commun. Image Represent.*, vol. 11, pp. 245–265, 2000.
- [21] P. Maragos, "Algebraic and PDE approaches for lattice scale-spaces with global constraints," *Int. J. Comput. Vis.*, vol. 52, no. 2–3, pp. 121–137, 2003.

- [22] L. Vincent, "Grayscale area openings and closings, their efficient implementations," presented at the 1st Workshop on Mathematical Morphology and its Applications to Signal Processing, 1993.
- [23] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Feb. 1993.
- [24] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998.
- [25] Y. Meyer, "Oscillating patterns in image processing and nonlinear evolution equations," ser. Univ. Lecture AMS vol. 22, 2002.
- [26] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math.*, vol. 42, no. 5, pp. 577–684, 1989.
- [27] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [28] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D: Nonlin. Phenomena*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.
- [29] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Sci. Comput.*, vol. 19, no. 1–3, pp. 553–572, 2003.
- [30] L. Alvarez, P.-L. Lions, and J.-M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM J. Numer. Anal.*, vol. 29, no. 3, pp. 845–866, 1992.
- [31] J.-F. Aujol and A. Chambolle, "Dual norms and image decomposition models," *Int. J. Comput. Vis.*, vol. 63, no. 1, pp. 85–104, June 2005.
- [32] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.
- [33] P. Maragos and A. C. Bovik, "Image demodulation using multidimensional energy separation," *J. Opt. Soc. Amer. A*, vol. 12, no. 9, pp. 1867–1876, 1995.
- [34] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "The multi-component AM-FM image representation," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 1094–1100, Jun. 1996.
- [35] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Multidimensional quasi-eigenfunction approximations and multi-component am-fm models," *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 227–242, Feb. 2000.
- [36] I. Kokkinos, G. Evangelopoulos, and P. Maragos, "Advances in texture analysis: Energy dominant components and multiple hypothesis testing," presented at the IEEE Int. Conf. Image Processing, 2004.
- [37] T. Lindeberg, "Features detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 79–116, 1998.
- [38] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [39] F. Meyer and P. Maragos, "Multiscale morphological segmentations based on watershed, flooding, and eikonal PDE," in *Proceedings of Scale-Space*. New York: Springer-Verlag, 1999, pp. 351–362.
- [40] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations," *J. Comput. Phys.*, vol. 79, pp. 12–49, 1988.
- [41] J. A. Sethian, *Level Set Methods and Fast Marching Methods*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [42] J. N. Tsitsiklis, "Efficient algorithms for globally optimized trajectories," *IEEE Trans. Autom. Control*, vol. 40, no. 9, pp. 1528–1538, Sep. 1995.
- [43] P. C. J. Rivest, "Localizing contours defined by more than one attribute," *Vis. Res.*, vol. 36, no. 1, pp. 53–66, 1996.
- [44] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *IJCV*, vol. 22, no. 1, pp. 61–79, 1997.
- [45] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 2, pp. 158–175, Feb. 1995.
- [46] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognit.*, vol. 29, no. 8, pp. 1335–1346, 1996.
- [47] J. Liu and Y. Yang, "Multiresolution color image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 7, pp. 689–700, Jul. 1994.
- [48] 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," in *Proc. 8th Int. Conf. Computer Vision*, Jul. 2001, vol. 2, pp. 416–423.



Anastasia Sofou received the B.Sc. degree from the Department of Informatics and Telecommunications, Kapodistrian University of Athens, Athens, Greece, in 1998, the M.Sc. degree in 1999 from the Computer Science Department, University of Bristol, U.K., and the Ph.D. degree in electrical and computer engineering from the National Technical University of Athens in 2006.

Her research interests include computer vision, pattern recognition, and digital signal processing.



Petros Maragos (S'81–M'85–SM'91–F'95) received the Diploma in electrical engineering from the National Technical University of Athens (NTUA), Athens, Greece, in 1980, and the M.Sc.E.E. and Ph.D. 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, Harvard University, Cambridge, MA, where he worked for eight years as a Professor of electrical engineering. In 1993, he joined the faculty of the ECE School, 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 a Professor at the School of Electrical and Computer Engineering, 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.

Dr. Maragos 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; as editorial board member for the journals *Signal Processing* and *Visual Communications and Image Representation*; as General Chairman or Co-Chair of conferences (VCIP'92, ISMM'96, VLBV'01, MMSP'07); and as a member of the IEEE DSP committees. His research has received several awards, including a 1987 National Science Foundation Presidential Young Investigator Award; the 1988 IEEE Signal Processing (SP) Society's Young Author Paper Award for the paper "Morphological Filters"; the 1994 IEEE SP Senior Award; the 1995 IEEE Baker Award for the paper "Energy Separation in Signal Modulations with Application to Speech Analysis"; the 1996 Pattern Recognition Society's Honorable Mention Award for the paper "Min-Max Classifiers"; and the 2007 EURASIP Technical Achievements Award for contributions to nonlinear signal processing and systems theory, image processing, and speech processing.