Color Texture Segmentation Based on the Modal Energy of Deformable Surfaces

Michail Krinidis and Ioannis Pitas, *Fellow, IEEE*

Abstract—This paper presents a new approach for the segmentation of color textured images, which is based on a novel energy function. The proposed energy function, which expresses the local smoothness of an image area, is derived by exploiting an intermediate step of modal analysis that is utilized in order to describe and analyze the deformations of a 3-D deformable surface model. The external forces that attract the 3-D deformable surface model combine the intensity of the image pixels with the spatial information of local image regions. The proposed image segmentation algorithm has two steps. First, a color quantization scheme, which is based on the node displacements of the deformable surface model, is utilized in order to decrease the number of colors in the image. Then, the proposed energy function is used as a criterion for a region growing algorithm. The final segmentation of the image is derived by a region merge approach. The proposed method was applied to the Berkeley segmentation database. The obtained results show good segmentation robustness, when compared to other state of the art image segmentation algorithms.

Index Terms—Color segmentation, color quantization, energy function, image segmentation, modal analysis, 3-D deformable models.

I. INTRODUCTION

IMAGE segmentation plays a fundamental role in many computer vision applications. In pattern recognition approaches [1], image segmentation enables the isolation of single objects or their parts in the scene that can be subsequently identified in an easier and more accurate way. A continuous research effort has been also made in automatic image annotation [2] and retrieval [3], where image segmentation is useful, since the aforementioned procedures are expensive and time consuming. Moreover, a number of studies based on image segmentation have been recently reported regarding scene analysis [4]. Image segmentation is also used as an initial step in object tracking [5], where the exact position of an object is crucial. In all the aforementioned applications, the main aim is to separate an image into homogeneous regions, a step that has been proved to boost their overall algorithmic performance.

Challenging problems in image segmentation arise from the existence of images depicting natural scenes at low resolution, uneven illumination, scale changes etc. The human visual system is very good at segmenting a natural scene image into perceptually homogeneous regions. However, it is extremely difficult to automatically segment such images into regions with perceptual meaning. Hence, in the last two decades, intensive research has been carried out in this area. The research interest in this paper is focused on unsupervised image segmentation algorithms, which can be broadly divided in three categories region-based, graph-based and feature-based ones [6], [7]. Most such image segmentation approaches rely on the global optimization of a segmentation criterion. Additional information on the aforementioned image segmentation categories can be found in review publications [8], [9].

Region-based image segmentation approaches [10]–[12] try to separate connected image regions by grouping neighboring pixels based on brightness, color, and texture. Afterwards, adjacent regions are merged, under some criterion involving homogeneity or region boundary sharpness. In [10], the combined use of statistical distributions of filter responses for texture characterization within a level set framework, is proposed. The segmentation is typically based on the minimization of a region-based functional. This functional is defined as a weighted Kullback–Leibler measure, which is based on the distributions of texture filter responses computed inside the image regions and on regularity constraints at the region boundaries.

Many research efforts have taken place regarding graph-based image segmentation algorithms [13]–[16]. These approaches use an undirected graph to represent the image pixels. Each graph edge has a weight representing pairwise pixel similarity. The objective is to minimize the cost of splitting the graph into a number of disjoint sets. In [13], each image pixel corresponds to a node in the graph and neighboring pixels are connected by undirected edges. The dissimilarity between pixels is measured and weights are assigned to each graph edge. The segmentation criterion is based on the degree of variability in neighboring regions of the image. Thus, any segmentation is induced by a subset of the graph edges. In [14], a graph-theoretic criterion called *normalized cut* was introduced for measuring the suitability/effectiveness of an image partition. The minimization of this criterion was formulated as a generalized eigenvalue problem. The eigenvectors were used to construct satisfactory partitions of the image and the process was repeated recursively in order to obtain the image regions.

The suitability of image features for image segmentation has been investigated by many researchers [17]–[19]. Color and texture features are exploited along with the knowledge

of human visual perception mechanisms in order to assess image region homogeneity. In [17], the mean-shift algorithm is used for feature space analysis, where the feature space can be regarded as the empirical probability density function of the represented parameter. The mean-shift is defined as the difference between feature samples and their weighted mean, by using a radially symmetric kernel for the weights and the sample points as the kernel center. The mean-shift method converges to a nearby point where the estimate of the density function has zero gradient. Thus, each image pixel is associated with a mode of the joint density located in its neighborhood, after the application of mean-shift space analysis. Afterwards, the pixels are grouped by a merging algorithm. In [7], 2-D texture filter banks or simple fixed-size filter windows were used in order to obtain texture features. The distribution of the texture features is modeled by using a mixture of Gaussian degenerate distributions. The mixture is effectively segmented by a simple agglomerative clustering algorithm derived from a lossy data compression approach. The image segmentation is achieved by minimizing the overall coding length of the feature vectors.

This paper presents an unsupervised color texture image segmentation algorithm, which will be subsequently called *modal image segmentation* (MIS) method. The MIS approach was mainly motivated by the technique presented in [20], which aims at selecting and tracking feature points on video sequences using deformable surface models. In [20], an intermediate result of the deformation procedure was shown to be a novel combination output of various line and edge detection masks. Thus, it can be appropriate for boundary detection of neighboring image regions. Based on this reasoning, we introduce a new deformable model energy function, which expresses the local smoothness of one image region, in order to segment color images. The core of this work consists of two steps. First, a color quantization scheme, which is based on the node displacements of the deformable surface model, is utilized in order to obtain a coarse image representation. Second, the proposed energy function is used as a criterion for image segmentation. A flow diagram, which depicts the different modules of the MIS algorithm, is shown in Fig. 1. The image intensity and the spatial image information are combined and act as external forces to a physics based deformable surface model. The proposed color quantization scheme is based on the node displacements of a deformable surface model (DSM) that is fit to the image. The image intensity is considered to act as the external force of a 3-D DSM. A weight, expressing the smoothness of the local image areas is assigned to each image pixel, which are subsequently separated into color clusters.

Modal analysis, which is used in order to describe and analyze the surface deformations, aims at evaluating the proposed energy function. It is an alternative approach to the direct integration of the system of differential equations that result from any finite element method [21]. It is based on mode superposition, i.e., expressing the time-dependent mode deformations as the superposition of the natural modes of the system. In the proposed method, the model deformation equations were expanded in order to achieve a fast deformation implementation. Modal
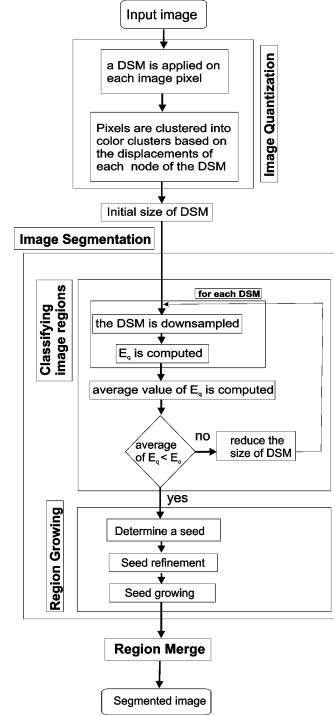

Fig. 1. Flow diagram of the proposed image segmentation algorithm.

analysis techniques were also used in a variety of different applications for solving model deformations, i.e., for analyzing nonrigid object motion [21], for the alignment of serially acquired slices [22], for multimodal brain image analysis [23], for image compression [24] and for 2-D object tracking [20]. In this paper, modal analysis is used in a totally different and novel way, i.e., for color texture image segmentation.

The main contribution of this paper essentially is the design of a novel energy criterion based on the use of local deformable surfaces involving modal analysis for image modelling. The local spatial information of image areas, which depends on the output of color quantization, is used as external forces that drive the DSM. Image segmentation is performed by combining the DSM energy with a region growing algorithm adapted to the proposed energy function. A final region merging algorithm provides the overall image segmentation. The details of the algorithm will be provided below. The MIS algorithm was tested on the Berkeley Segmentation database [25], which consists of a subset of the Corel image dataset. The results indicate that MIS algorithm outperforms other well known image segmentation algorithms.

Summarizing, the novel contributions of this study are: A new energy function, derived from modal analysis techniques, is used as a criterion for color texture image segmentation. This energy function employs a byproduct of the deformation procedure to express the local smoothness of an image region. A combination of the spatial information with the image intensity, is used in a novel way, as external forces that drive the DSM. A way to tailor the model deformation equations in order to achieve a fast implementation of the node deformations. A color quantization scheme which exploits the node displacements of

the 3-D DSM. A region growing procedure [12], which minimizes the proposed energy function.

The remainder of the paper is organized as follows. In Section II, a brief description of the deformation procedure based on modal analysis and the derivation of the proposed energy function are presented, along with a fast solution of the deformation equations. The color quantization method is described in Section III. In Section IV, the image segmentation algorithm is introduced. The performance of the proposed technique is studied in Section V. Finally, conclusions are drawn in Section VI.

II. Modal Analysis of a Deformable Surface Model and Derivation of the Modal Energy Function

In this section, the physics-based DSM that is used along with modal analysis [24], will be briefly reviewed and a new energy function derived from modal analysis will be introduced. Furthermore, a fast solution of the deformation process, based on modal analysis, will be described.

The image intensity $I(x, y)$ can be assumed to define a surface over the image domain (x, y) that will be subsequently called *intensity surface*. The MIS approach focuses on parameterizing this surface defined by $(x, y, I(x, y))$, also called XYI space [26]. A 3-D physics-based deformable surface model (DSM), introduced in [21], [23], and [27], is used for this purpose. The DSM consists of a uniform quadrilateral mesh of $N = N_h \times N_w$ nodes, as illustrated in Fig. 2. In this section, we assume that N_h, N_w are equal to the image height and width respectively (in pixels), so that each image pixel corresponds to one mesh node. Each model node is assumed to have a mass m and is connected to its neighbors with perfect identical springs of stiffness k, having natural length l_0 and damping coefficient c. Under the influence of internal and external forces, the mass-spring system is deformed to provide a 3-D mesh representation of the image intensity surface.

In our case, the initial and the final deformable surface states are known. The initial state is the initial model configuration and the final state is the image intensity surface. Therefore, it can be assumed that a constant force load $\mathbf{f}$ is applied to the surface model [23]. Thus, the physics-based model is governed by the static problem formulation

$$\mathbf{K}\mathbf{u} = \mathbf{f} \qquad (1)$$

where $\mathbf{K}$ is the $N \times N$ stiffness matrix and $\mathbf{u}$ is the nodal displacement vector. Instead of finding directly the equilibrium solution of (1), one can transform it by a basis change [28]

$$\mathbf{u} = \mathbf{\Phi}\tilde{\mathbf{u}} = \sum_{\iota=1}^{N=N_h N_w} \boldsymbol{\phi}_\iota \tilde{\mathbf{u}}_\iota \qquad (2)$$

where $\tilde{\mathbf{u}}$ is referred to as the *generalized displacement* vector, $\tilde{\mathbf{u}}_\iota$ is the ιth component of $\tilde{\mathbf{u}}$ and $\mathbf{\Phi}$ is a matrix of order N, whose columns are the eigenvectors $\boldsymbol{\phi}_\iota$ of the generalized eigenproblem

$$\mathbf{K}\boldsymbol{\phi}_\iota = \omega_\iota^2 \mathbf{M}\boldsymbol{\phi}_\iota \qquad (3)$$

where $\mathbf{M}$ is the mass matrix of the deformable model, whose formulation is explained analytically in [27]. The ιth eigenvector $\boldsymbol{\phi}_\iota$, i.e., the ιth column of $\mathbf{\Phi}$, is also called the ιth *vibration mode* and ω_ι is the corresponding eigenvalue (also called *vibration frequency*). Equation (2) is known as *modal superposition equation*. A significant advantage of this formulation is that the vibration modes (eigenvectors) $\boldsymbol{\phi}_\iota$ and the frequencies (eigenvalues) ω_ι of a plane topology have an explicit formulation [21]

$$\omega^2(i, j) = \frac{4k}{m}\left[\sin^2\left(\frac{\pi i}{2N_h}\right) + \sin^2\left(\frac{\pi j}{2N_w}\right)\right] \qquad (4)$$

$$\phi_{n,n'}(i, j) = \cos\frac{\pi i(2n-1)}{N_h}\cos\frac{\pi j(2n'-1)}{N_w} \qquad (5)$$

where $i = 0, 1, \ldots, N_h - 1, j = 0, 1, \ldots, N_w - 1, n = 1, 2, \ldots, N_h, n' = 1, 2, \ldots, N_w, \omega^2(i, j) = \omega_{i N_w + j}^2, \phi_{n,n'}(i, j)$ is the (n, n')th element of matrix $\boldsymbol{\phi}(i, j) = \boldsymbol{\phi}_{i N_w + j}$. Thus, they do not have to be computed using eigen-decomposition techniques. In the modal space, (1) can be written as

$$\tilde{\mathbf{K}}\tilde{\mathbf{u}} = \tilde{\mathbf{f}} \qquad (6)$$

where $\tilde{\mathbf{K}} = \mathbf{\Phi}^T \mathbf{K}\mathbf{\Phi}$ and $\tilde{\mathbf{f}} = \mathbf{\Phi}^T \mathbf{f}, \mathbf{f}$ being the external force vector. Hence, by using (2), (4), and (5), (7) is simplified to $3N$ scalar equations

$$\omega_\iota^2 \tilde{u}_\iota(r) = \tilde{f}_\iota(r) \qquad (7)$$

where $r = x, y, z, \tilde{u}_\iota(r)$ is the rth component of the ιth vector of $\tilde{\mathbf{u}}$ and $\tilde{f}_\iota(r)$ is the rth component of the external force vector $\tilde{\mathbf{f}}$ acting on node ι.

In our case, the image intensity of interest is described in terms of the vibrations of an initial model. The size of the model (in number of nodes) that was used to parameterize the image surface was equal to the image size (in pixels). The quadrilateral mesh model is initialized and the elements $\tilde{u}_{iN_w+j}(r)$ are explicitly computed by using (2), (4), (5), and (7), as follows:

$$\tilde{u}_{iN_w+j}(r) = \frac{\sum_{n=1}^{N_h}\sum_{n'=1}^{N_w} F_r(n, n')\phi_{n,n'}(i, j)}{(1 + \omega^2(i, j))\sqrt{\sum_{n=1}^{N_h}\sum_{n'=1}^{N_w}\phi_{n,n'}^2(i, j)}} \qquad (8)$$

where $F_r(n, n')$ is the rth component of the external force vector acting on pixel (n, n'), i.e., $F_r(n, n') = f_{nN_h+n'}(r)$.

A new energy function, called *modal energy* (ME), is introduced by using the elements $\tilde{u}_{i N_w+j}(r)$ of the DSM [see (9), shown at the bottom of the next page], and ME is given by E_q. The scalar value E_q expresses the local smoothness of an image region. Its usefulness for image segmentation will be shown in Section IV.

By using (2) and (8), it can be found that the deformations $u_{xy}(r)$ along the rth axis of the deformable model node that corresponds to pixel (x, y), based on the modal analysis for a plane topology, can be described by (10), shown at the bottom
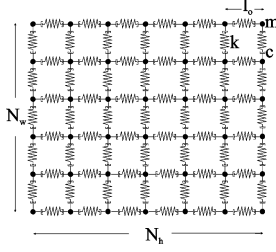
Fig. 2. Quadrilateral image intensity surface (mesh) model.

of the next page, where $u_{xy}(r)$ is the r component of the element $\mathbf{u}_{(x-1)N_w+y}$ of vector $\mathbf{u}$ and $r = x, y, z$. The above equation presents the final outcome of the aforementioned methodology of deformable surface models adaptation based on Modal Analysis. The deformation-based surface modeling approach has heavy computational burden, especially when real time applications are concerned. Hence, we present a fast implementation of the deformation process, which is based on the discrete modal transform (DMT) presented in [24]. The components of the deformations $\mathbf{u}$ of the 3-D DSM applied on an image intensity $I(x, y)$ along each axis are given by (10), which can be rewritten as

$$u_{xy}(r) = \sum_{i=0}^{N_h-1} \sum_{j=0}^{N_w-1} \mathcal{F}(i,j)$$
$$\times \frac{\phi_{x,y}(i,j)}{\sqrt{\sum_{n=1}^{N_h} \sum_{n'=1}^{N_w} \phi_{n,n'}^2(i,j)}},$$
$$x = 1, \ldots, N_h, \quad y = 1, \ldots, N_w \quad (11)$$

and

$$\mathcal{F}(i,j) = \frac{1}{\left[1 + \lambda \left(\sin^2\left(\frac{\pi i}{2N_h}\right) + \sin^2\left(\frac{\pi j}{2N_w}\right)\right)\right]}$$
$$\times \mathcal{C}(i,j) \quad (12)$$

where $\mathcal{F}(i,j)$ is the DMT, $\mathcal{C}(i,j)$ is the 2-D discrete cosine transform (DCT), which is defined in [29] and $\lambda = 4$ k/m. In a similar manner, the elements $\tilde{u}_{i\,N_w+j}(r)$ can be rewritten by using DMT as follows:

$$\tilde{u}_{i\,N_w+j}(r) = \mathcal{F}(i,j). \quad (13)$$

One can easily see that the DMT corresponds to a nonseparable image transform. Thus, its computational complexity for a $N \times N$ image is of the order of $O(N^4)$. However, since

2-D DMT can be calculated using DCT through (12), its complexity can be reduced to $O(N^2 \log_2 N)$, if fast DCT implementations are used [30]. For the experiments performed in this paper, the computations of the *generalized displacement* vector $\tilde{u}_{i\,N_w+j}(r)$ and the deformations $u_{xy}(r)$ along the rth axis of the DSM were computed using (13) and (11), respectively.

III. COLOR QUANTIZATION

In this section, the first step of the overall image segmentation method consisting of a color quantization algorithm is briefly described. It provides a few representative colors that can describe different image regions. Its basic idea comes from the peer group filtering method, which was introduced in [31]. The original algorithm was extended, so as the feature vector of the peer group filtering includes the DSM node displacement (11), instead of the color information of each pixel of the image.

A DSM $\mathbf{P}$ of size $R_w \times R_w$ is applied on pixel $\mathbf{x}_q = (x_q, y_q)$ of image I. The components of the external forces $\mathbf{f}$ in (1) acting on the DSM are taken to be equal to zero along the x and y axes, i.e., $f_i(x) = f_i(y) = 0$ and $i = 0, \ldots, R_w^2$. On the other hand, the components of these forces along the z (intensity) axis are taken to be proportional to the pixel intensity (x_q, y_q): $f_{(x_q-1)N_w+y_q}(z_q) = I(x_q, y_q)$, where $f_{(x_q-1)N_w+y_q}(z_q)$ is the component along the z axis of the $(x_q - 1)N_w + y_q$th element of vector $\mathbf{f}$. Under such a condition, the model is deformed only along the z axis, representing image intensity. Three different color components act on each pixel of the image and the deformations are computed independently for the color components. For each DSM node, the following measure is computed:

$$d_{iq} = \|\mathbf{u}_q - \mathbf{u}_i\| \quad (14)$$

where $\mathbf{u}_i$ stores the displacement of the ith node of the DSM for all the color components and $\|\cdot\|$ is the Euclidean distance. By sorting all the pixels $\mathbf{x}_i$ in the window W centered around the pixel $\mathbf{x}_q$, according to their distances d_{iq} in ascending order, one can construct the following feature vector $\mathbf{G}_q$, which is called peer group for pixel $\mathbf{x}_q$

$$\mathbf{G}_q = \{\mathbf{x}_i, \quad i \in [1, R_w^2]\} \quad (15)$$

where $\mathbf{x}_i$ is the ith pixel in the window W. The appropriate cardinality of each peer group is defined by using the distances d_{iq} for Fisher's discriminant estimation, as described in [31]. The peer group cardinality can be considered as a parameter for the control of the noise removal and the smoothing that peer group filtering causes to the image. By applying a DSM of size

$$E_q = \frac{1}{\frac{1}{N_h N_w - 1} \sqrt{\sum_{i=1}^{N_h} \sum_{j=1}^{N_w} \left(\tilde{u}_{iN_w+j}(x_{iN_w+j}) + \tilde{u}_{iN_w+j}(y_{iN_w+j})\right)^2}} \quad (9)$$

$$u_{xy}(r) = \sum_{i=0}^{N_h-1} \sum_{j=0}^{N_w-1} \frac{\sum_{n=1}^{N_h} \sum_{n'=1}^{N_w} F_r(n,n') \phi_{n,n'}(i,j)}{(1 + \omega^2(i,j)) \sum_{n=1}^{N_h} \sum_{n'=1}^{N_w} \phi_{n,n'}^2(i,j)} \phi_{x,y}(i,j) \quad (10)$$
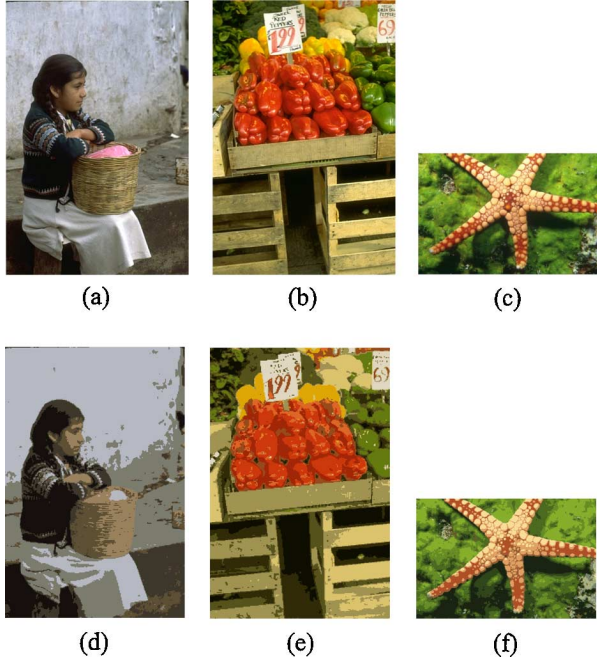
Fig. 3. (a)–(c) Initial color images. (d)–(f) Corresponding quantized images.

$R_w \times R_w$ to each pixel $\mathbf{x}_q$ of image I, every pixel of the image is associated with one weight that indicates the smoothness of the area around the pixel. The weight $g(\mathbf{x}_q)$ is defined as

$$g(\mathbf{x}_q) = e^{-T_q} \tag{16}$$

where T_q is the maximum distance d_{iq} in $\mathbf{G}_q$. Thus, pixels in smoothed areas are associated with bigger weights than pixels in textured areas. We use this weight to perform color image quantization.

A modified version of the general Lloyd algorithm [32] was used to perform color image quantization. Let us assume that the image I is quantized into a number of C_n color clusters. The initial value of C_n is set to be twice as great as the average value of $g(\mathbf{x}_q)$ for the entire image. The weighted distortion measure D_n of the color cluster C_n is given by

$$D_n = \sum_{\mathbf{x}_i \in C_n} g(\mathbf{x}_i)\|\mathbf{x}_i - \mathbf{c}_n\|^2 \tag{17}$$

where $\mathbf{c}_n$ is defined as

$$\mathbf{c}_n = \frac{\sum_{\mathbf{x}(i) \in C_n} g(\mathbf{x}_i)\mathbf{x}(i)}{\sum_{\mathbf{x}(i) \in C_n} g(\mathbf{x}_i)}. \tag{18}$$

The distortion measure is used to detect which clusters should be split in order to reach the number of the initial clusters. The use of the weights $g(\mathbf{x}_i)$ in (17) ensures that pixels in textured image areas will not influence color quantization as much as pixels in homogeneous image pixels do.

In the final step, all clusters, whose minimum distance between their centroids is less than a predefined threshold, are merged by exploiting an agglomerative clustering algorithm [33]. The quantized image is obtained by assigning to each pixel its closest cluster centroid. An example of the proposed color image quantization algorithm is depicted in Fig. 3. The

first image [Fig. 3(a)] was separated into seven color clusters, the second image [Fig. 3(b)] in 11 color clusters and the third one [Fig. 3(c)] in 12 color clusters.

IV. IMAGE SEGMENTATION BASED ON THE MODAL ENERGY

In this section, a color image segmentation algorithm that combines color quantization and 3-D deformable surface models, is introduced. The color quantization algorithm described in the previous section, is applied to the image in order to extract a coarse image representation. Thus, the intensity of the image in this section, is the output of the quantization procedure.

Let us assume that a 3-D DSM of size $R_W \times R_W$ is applied to a feature point $\mathbf{x}_q = (x_q, y_q)$ of the image I. In this section, the components of the external forces $\mathbf{f}$ in (1) acting on the DSM are taken to be equal to zero along the z (image intensity) axis, since they were exploited in the color quantization procedure. In the image segmentation module, the spatial information of the DSM is exploited, i.e., the x and y components of the DSM at the corresponding axes. Thus, the components of the external forces along the x axis are evaluated as follows:

$$f_{i,x} = \begin{cases} x_i - \bar{x}, & I(x_i) = I(x_q) \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

where $\bar{x}$ is the center of gravity of all the pixels $\mathbf{x}_i$ that have the same intensity with $\mathbf{x}_q$. In a similar manner, the components of the external forces along the y axis are taken to be equal to the spatial information of y axis of the DSM

$$f_{i,y} = \begin{cases} y_i - \bar{y}, & I(x_i) = I(x_q) \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

where $\bar{y}$ is the center of gravity of all the pixels $\mathbf{x}_i$ that have the same intensity with $\mathbf{x}_q$.

To achieve image segmentation, the MIS approach applies on each image pixel $\mathbf{x}_q$ a 3-D DSM of size $R_W \times R_W$ and subsequently exploits the proposed energy function E_q for image pixel $\mathbf{x}_q$ by using (9). The scalar value E_q expresses the local smoothness of one image region, i.e., the lower the E_q value is, the more likely the image region around the pixel $\mathbf{x}_q$ corresponds to a smooth region. Hence, E_q can be considered to be an indicator of whether a pixel is close to a region boundary or not. In Fig. 4, one can see the ME E_q for all the pixels of two different images. The size of the deformable model was 50×50 nodes, applied to images consisting of 481×321 pixels. The image in Fig. 4(a) has rather uniform image regions, when compared to the image of Fig. 4(c), which contains more details. The average E_q value for the image in Fig. 4(a) is 0.3807, while the image in Fig. 4(c) has almost the double value (0.6855), since the latter has more pixels lying on region boundaries.

The size $R_w \times R_w$ of the DSM determines the size of the image region that is examined. Since it is difficult to know the proper size $R_w \times R_w$ of the DSM *a priori*, we propose using an adaptive window size, defined as the *best region* for detecting the region boundaries in image I. Small deformable surfaces (i.e., small window size $R_w \times R_w$) are useful for tracing boundaries of neighboring regions, while large deformable surfaces (i.e., large window size $R_w \times R_w$) are appropriate for detecting texture boundaries. Starting with a large size for the
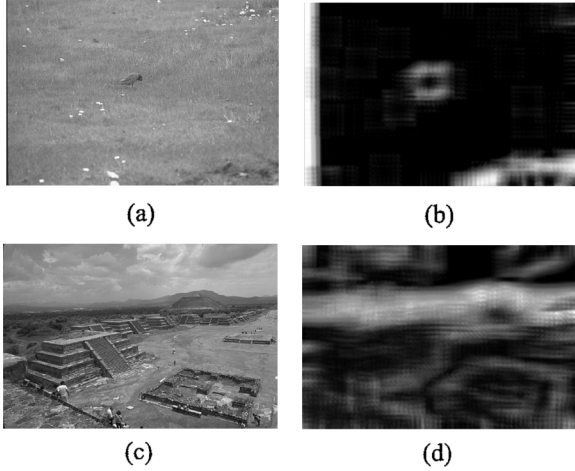
Fig. 4. Example of the ME energy E_q for two different images from Berkeley database. (a) An image with uniformly image intensity and (b) its energy term E_q. (c) An image with several different image regions and contours and (d) its energy term E_q.
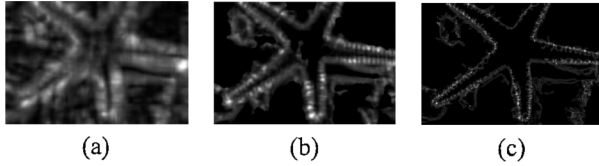

Fig. 5. Example of the adaptive evaluation of the deformable model size, applied to image of Fig. 3(c). (a) The ME energy term E_q for all the image pixels with model size 100×100 at iteration 1 (average value of $E_q = 0.2676$). (b) The ME energy term E_q for all the image pixels with model size 50×50 at iteration 2 (average value of $E_q = 0.1361$). (c) The ME energy term E_q for all the image pixels with model size 30×30 at iteration 3 (average value of $E_q = 0.0001$), where the algorithm converges.

DSM, which depends on the initial size of the image I (e.g., a 512×512 DSM size for a 1024×756 image), a DSM is applied to each image pixel. Then, the algorithm iteratively decreases the DSM size down to a certain limit (the smallest size used was 8×8), until the average value of E_q (for the entire image) surpasses a threshold E_o. Hence, the MIS algorithm starts with a crude segmentation of the image of interest, where the textured image areas are localized. At each iteration, where the DSM size is decreased, the border refinement of neighboring regions is achieved. When a DSM exceeds the image borders, the first/last row or column at the image border is repeated and constitutes the external forces of the DSM. For the evaluation of a suitable value for E_o, E_q values were computed for different sizes of the deformable model for different images of interest. These experiments have demonstrated that one can achieve a good indication of the image region smoothness by setting the threshold E_o equal to 0.01. For computational reasons, successive deformable surface models were downsampled at appropriate levels at each iteration. The final outcome is an image with intensities being the E_q values, which resulted from the aforementioned procedure. An example of the aforementioned procedure is depicted in Fig. 5.

Once all image pixels $\mathbf{x}_q$ have been assigned a value E_q, a region-growing algorithm, which is a modification of [12], is applied. The seed growing method is based on the ME values E_q
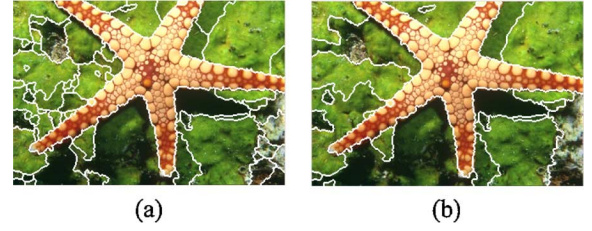

Fig. 6. (a) Initial segmentation of image in Fig. 3(c) after the application of seed growing algorithm and (b) the final segmentation after merging the over-segmented areas.

of each image pixel. The number of the different image regions depends on the values of E_q. First, the average value μ_{E_q} and the standard deviation σ_{E_q} of E_q of the image are computed and a threshold T

$$T = \mu_{E_q} + \alpha\sigma_{E_q} \qquad (21)$$

is derived. The scalar value α was set equal to 0.7, since this value provided good results in the experiments. Then, candidate seed points are considered to be all pixels that have $E_q < T$. The candidate seed points are connected based on the four-point connectivity, in order to grow image areas. Finally, they are considered to be image regions, only if their size R_N satisfies the following property:

$$R_N \geq \frac{R_W}{2}s \qquad (22)$$

where s is the sampling factor that was used to downsample the deformable model (of size $R_W \times R_W$) in the computation of E_q.

When the seeds have been determined, the region growing procedure provides the candidate image regions. First, the "holes" inside the seeds are removed ("hole" is a quite small image region located inside a big one). Then, the average value μ_{E_q} of all the unsegmented pixels (pixels that do not belong to any seed) is computed and the pixels with $E_q < \mu_{E_q}$ are connected to form growing areas. If a growing area is adjacent to only one seed, it is assigned to that seed. This procedure is repeated with smaller sizes of DSM in order to locate more accurately the region boundaries. Finally, the ME energy E_q is computed for all the remained unclassified pixels and the ones with the minimum E_q values are connected to their adjacent seeds. The final step is repeated until all the pixels belong to an image region. An example of the seed growing algorithm on a color image is depicted in Fig. 6(a).

An agglomerative merge algorithm [33], follows the region-growing procedure, since the segmentation procedure may result in oversegmentation. The merging approach is based on the color similarities of neighboring areas. The Euclidean distance of color histograms between any two neighboring image areas is computed and stored in a matrix E. The mean value and the variance of this matrix E are estimated. The pair of regions with the minimum Euclidean distance is merged and the procedure is repeated up to a maximum threshold of Euclidean distance, which is equal to the mean value of the matrix E minus its variance. The uniform CIE LUV color space is used. An example of

the merging algorithm on a color image is depicted in Fig. 6(b). The interaction among the proposed quantization scheme, the ME energy, the region growing algorithm and the merging algorithm was experimentally studied. A set of experiments is described in detail in Section V.

V. PERFORMANCE EVALUATION

A. Evaluation Setup and Dataset

Comprehensive experiments were conducted in natural scene images to evaluate the performance of the MIS method. The proposed method has been used to segment an image into distinct color-textured regions on the Berkeley segmentation database [25]. This database was selected because it contains hand-labeled segmentations of the images from 30 human subjects. Half of the segmentations involve color images and the other half grayscale images. The database comprises of various images from the Corel dataset and contains ground truth of 300 images for benchmarking image segmentation and boundary detection algorithms. The content of the images are landscapes, animals, portraits and various objects. The proposed algorithm was applied to all 300 images and the output was compared to human perceptual ground truth.

The metrics used for the quantitative evaluation of the proposed algorithm were the following.

- The *Probabilistic Rand index* (PR) [34] allows the comparison of a test segmentation algorithm using multiple ground truth images. It measures the fraction of pixel pairs, whose labels are consistent in the test segmentation and the ground truth one. The PR averages over multiple ground truth segmentations and takes values in the $[0, 1]$ interval, where 0 means that the acquired segmentation has no similarities with the ground truth and 1 means that the test and ground truth segmentations are identical.
- The *Normalized Probabilistic Rand index* (NPR) [35] is an extension of the PR and is normalized with the expected value of the PR over the test data. This fact makes the NPR more sensitive while having a larger range.
- The *Global Consistency error* (GCE) [25] measures the tolerance up to which one image segmentation can be considered as a refinement of another one. A zero value of GCE means that the segmentations are identical, whereas a nonzero value means that inconsistency exists between the two image segmentations.
- The *Variation of Information* (VI) [36] measures the amount of information that is lost or gained from one image cluster to another one. The values of VI are always positive and zero value means that all the image clusters are identical.
- The *Boundary Displacement error* (BDE) [8] is the average displacement error between the boundaries of the extracted image regions. Lower values of BDE indicate high image segmentation quality.

B. Experimental Results

The first set of experiments dealt with the evaluation of the correctness of the segmentations produced by MIS with a varying set of parameters. The factor λ in the denominator
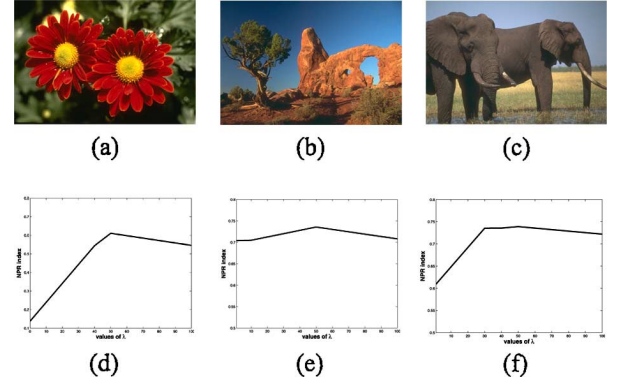

(a) (b) (c)

(d) (e) (f)

Fig. 7. (a)–(c) Original images and (d)–(f) the corresponding NPR index for different values of λ.
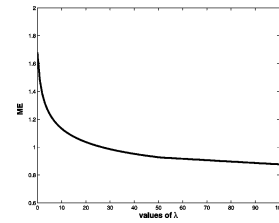

Fig. 8. ME E_q for different values of λ for a DSM applied to a uniform image region.

of (12) can be used to achieve multiple segmentations of an image. The NPR was computed for different values of λ for several images of the Berkeley database. Fig. 7 depicts the NPR index versus different values of λ. The NPR index increases, as λ increases up to the value of 50. Hence, an appropriate value of λ can be set equal to 50. This fact is also illustrated by Fig. 8, where ME E_q is depicted for different values of the factor λ for a DSM applied to a uniform image region. The ME E_q maintains a constant value for values of $\lambda \geq 50$, which is an another indication that the segmentation performance is constant for a rather wide range of λ values. One can say that the λ factor in (12) essentially acts as a parameter which tends to oversegment the image, when its value decreases. Some segmentation results for different values of λ are shown in Fig. 9. Again, from quantitative point of view, no great sensitivity is shown versus changes in λ for values equal to 50 and 100.

In order to evaluate the computational complexity of the MIS algorithm, the average computational time was computed for a test set of images (20 images selected randomly from the Berkeley database). The computational time was computed for two versions of the MIS algorithm. In the first version (denoted by MIS-Sl), the original deformation equations were exploited in order to compute ME, while in the second one, a fast implementation of the deformation procedure was used, as described in Section II. All the experiments were performed on an Intel Pentium 4 (3.01 GHz) processor PC with 1.5 GB of RAM for 317×211 and 512×512 color images. In terms of computational complexity for an image of range 317×211 pixels, 73.2 s per image are required for MIS-Sl algorithm, while 2.6 s per image are required for the proposed MIS algorithm. Furthermore, for an image of range 512×512 pixels, 198.6 s per image
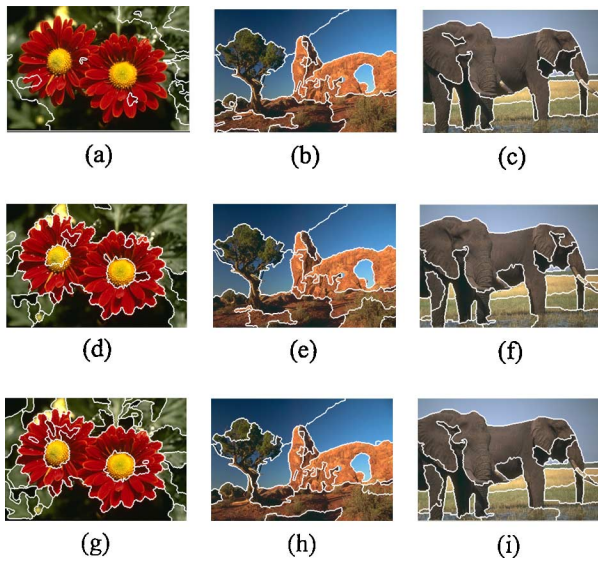
Fig. 9. (a)–(c) Image segmentations for $\lambda = 0$, (d)–(f) image segmentations for $\lambda = 50$, and (g)–(i) image segmentations for $\lambda = 100$.

TABLE I
AVERAGE VALUES FOR THE PR, BDE, VI, AND GCE ERRORS ACHIEVED ON 100 COLOR IMAGES OF THE BERKELEY SEGMENTATION DATABASE BY FOUR VARIANTS OF MIS ALGORITHM

Algorithms	Average Values			
	PR	BDE	VI	GCE
MIS	0.7712	7.6652	2.0042	0.2020
Kmeans-MIS	0.7281	8.1214	2.2602	0.2298
Q-MIS	0.7312	7.9038	2.1642	0.2141
NRM-MIS	0.7089	8.3497	2.3641	0.2484

are required for MIS-Sl algorithm, while 2.9 s per image are required for the proposed MIS algorithm. Thus, the computational speed of the MIS algorithm is rather good for segmenting images of size in the range of 512×512 pixels and it can be considered to be appropriate for offline processing of large images.

The aim of the third set of experiments was to find the best performance between three different variants of the proposed segmentation algorithm. In the first two variants, the proposed quantization scheme was replaced by two different approaches. In the first approach (denoted by K-means-MIS), the well known K-means algorithm [37] was exploited in order to classify the image colors into a predefined number of color clusters. The RGB color space was used, since this was also used for MIS. The number of color clusters was set to be equal to the one computed by the proposed quantization scheme, in order to achieve a fair comparison. In the second variant, the quantization scheme presented in [31] was used (denoted by Q-MIS), in order to reduce the number of the image colors. In the third variant (denoted by NRM-MIS), the image segments resulting from MIS were not merged. The Kmeans-MIS, Q-MIS, NRM-MIS, and MIS algorithms were tested on 100 color images, selected randomly from the Berkeley database and the results are summarized in Table I. In all the experiments, the parameters used for the four algorithms remained the same. Table I presents the average values for the PR, BDE, VI, and GCE errors for the four variants of MIS algorithm. One can see
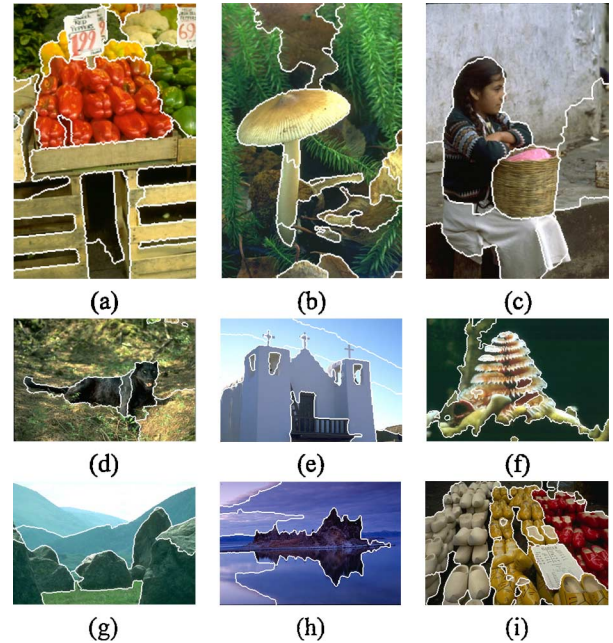

Fig. 10. Image segmentation based on the MIS algorithm. The factor λ in denominator (12) is equal to 50. The images are part of the Berkeley database.

TABLE II
AVERAGE VALUES FOR THE PR, BDE, VI, AND GCE ERRORS ACHIEVED ON THE BERKELEY SEGMENTATION DATABASE BY FIVE ALGORITHMS

Algorithms	Average Values			
	PR	BDE	VI	GCE
Ground truth	0.8754	4.9940	1.1040	0.0797
$MIS_{\lambda=0}$	0.7869	8.0930	2.0117	0.2110
$MIS_{\lambda=50}$	0.7981	7.8263	1.9348	0.1942
$MIS_{\lambda=100}$	0.7912	7.9213	1.9801	0.2022
Mean-Shift	0.7550	9.7001	2.4770	0.2598
NCuts	0.7229	9.6038	2.9329	0.2182
NNG	0.7841	9.9497	2.6647	0.1895
$CTM_{\gamma=0.1}$	0.7561	9.4211	2.4640	0.1767
$CTM_{\gamma=0.15}$	0.7627	9.4902	2.2035	0.1846
$CTM_{\gamma=0.2}$	0.7617	9.8962	2.0236	0.1877

that the MIS algorithm achieves the best performance. These experimental results provide evidence regarding the interaction between the proposed quantization scheme, the ME energy and the region growing algorithm, which is based on ME.

The next set of experiments aimed at testing the performance of the MIS algorithm and comparing it to that of well known image segmentation algorithms. In order to achieve unsupervised segmentation, the parameter λ of the DSM was set equal to 0, 50, and 100. Table II presents the average values for the PR, BDE, VI, and GCE errors on the Berkeley database for five unsupervised image segmentation algorithms. The values are averaged over the entire set of (300) images of the Berkeley database. The proposed MIS algorithm was compared to compression-based texture merging (CTM) [7], to mean-shift [17], to normalized cuts (Ncuts) [14] and to nearest neighbor graphs (NNG) [13] image segmentation algorithms. The results for the four algorithms were obtained from [7] and [9]. One can see that the proposed algorithm achieves better results than the other image segmentation algorithms for three (out of four) quantitative segmentation metrics for the three different values of λ. This fact provides quantitative evidence that MIS shows good

consistency for varying values of the parameter λ. More analytically, one can see that the average BDE error for the proposed algorithm is 7.8263, which is a large improvement over the error of 9.4211 achieved by CTM in [7] (almost 17% increase in segmentation accuracy). The BDE and GCE errors penalize the undersegmentation problem. The PR and VI errors seem to be more correlated with the ground truth segmentations. Thus, they can be considered to be objective indicators of the image segmentation performance. The MIS achieves the best average value of PR and VI errors on the Berkeley database. The variances of PR, BDE, VI, and GCE errors over the entire database for the MIS algorithm ($\lambda = 50$) are equal to 0.0234, 0.0291, 0.0264, and 0.0253, respectively, which also provide information regarding the robustness of the MIS algorithm performance. Some image segmentation results of the proposed algorithm are presented in Fig. 10.

VI. CONCLUSION

A new approach for color texture image segmentation based on 3-D deformable surface models is introduced in this paper. The proposed method combines knowledge of the intensity surface of the image and the spatial information of image regions, which subsequently drives a 3-D physics-based deformable surface model. The deformations of the deformable surface model are acquired by exploiting modal analysis techniques. A way to take advantage of the benefits of the model deformation equations in order to achieve a fast implementation of deformations, are also presented. The proposed image segmentation is based on two steps. First, color quantization, based on node displacements of the deformable surface model, is used to decrease the initial number of image colors. Then, a novel energy function, which is derived by exploiting an intermediate step of the deformation procedure, is assigned to each image pixel and expresses the smoothness of the image region around this pixel. The proposed energy function is used as an input to a region growing algorithm, in order to provide the image regions. Finally, a merging algorithm is exploited to acquire the final image segmentation. Results obtained on the Berkeley segmentation database indicate that the proposed algorithm achieves better quantitative results than four known unsupervised image segmentation algorithms.

REFERENCES

[1] J. C. Bezdek, L. O. Hall, and L. P. Clarke, "Review of MR image segmentation techniques using pattern recognition," *Med. Phys.*, vol. 20, no. 4, pp. 1033–1048, Jul. 1993.

[2] T. Athanasiadis, Y. Avrithis, and S. Kollias, "A semantic region growing approach in image segmentation and annotation," presented at the 1st Int. Workshop on Semantic Web Annotations for Multimedia, Edinburgh, U.K., May 2006.

[3] J. Puzicha, T. Hofmann, and J. Buhmann, "Histogram clustering for unsupervised segmentation and image retrieval," *Pattern Recognit. Lett.*, vol. 20, no. 9, pp. 899–909, Sep. 1999.

[4] M. Sharma, "Performance Evaluation of Image Segmentation and Texture Extraction Methods in Scene Analysis," Ph.D. dissertation, Univ. Exeter, U.K., 2000.

[5] K. Yamaoka, T. Morimoto, H. Adachi, T. Koide, and H. J. Mattausch, "Image segmentation and pattern matching based FPGA/ASIC implementation architecture of real-time object tracking," in *Proc. Conf. Asia South Pacific Design Automation*, New York, 2006, pp. 176–181.

[6] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1382–1389, Oct. 2007.

[7] A. Y. Yang, J. Wright, Y. Ma, and S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image Understand.*, 2007, to be published.

[8] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi, "Yet another survey on image segmentation: Region and boundary information integration," in *Proc. 7th Eur. Conf. Computer Vision III*, Copenhagen, Denmark, May 2002, vol. 2352/2002, pp. 21–25.

[9] A. Y. Yang, J. Wright, S. Sastry, and Y. Ma, Unsupervised Segmentation of Natural Images via Lossy Data Compression Univ. California, Berkeley, Tech. Rep. UCB/EECS-2006-195, Dec. 2006 [Online]. Available: http://chess.eecs.berkeley.edu/pubs/220.html

[10] I. Karoui, R. Fablet, J.-M. Boucher, and J.-M. Augustin, "Region-based image segmentation using texture statistics and level-set methods," presented at the IEEE Int. Conf. Acoustics, Speech and Signal Processing May 2006.

[11] T. Leung and J. Malik, "Contour continuity in region based image segmentation," *Lecture Notes Comput. Sci.*, vol. 1406, pp. 544–559, Jun. 1998.

[12] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 800–810, Aug. 2001.

[13] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sept. 2004.

[14] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 603–619, Aug. 2000.

[15] O. Rotem, H. Greenspan, and J. Goldberger, "Combining region and edge cues for image segmentation in a probabilistic gaussian mixture framework," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Minneapolis, MN, Jun/ 2007, pp. 1–8.

[16] C. Alzate and J. Suykens, "Image segmentation using a weighted kernel PCA approach to spectral clustering," in *Proc. IEEE Symp. Computational Intelligence in Image and Signal Processing*, Apr. 2007, pp. 208–213.

[17] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[18] J. Chen, T. Pappas, A. Mojsilovic, and B. Rogowitz, "Adaptive perceptual color-texture image segmentation," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1524–1536, Oct. 2005.

[19] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy data coding and compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, Sep. 2007.

[20] M. Krinidis, N. Nikolaidis, and I. Pitas, "2D feature point selection and tracking using 3D physics-based deformable surfaces," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 876–888, Jul. 2007.

[21] C. Nastar and N. Ayache, "Frequency-based nonrigid motion analysis: Application to four dimensional medical images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 11, pp. 1069–1079, Nov. 1996.

[22] S. Krinidis, C. Nikou, and I. Pitas, "Reconstruction of serially acquired slices using physics-based modelling," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 4, pp. 394–403, Dec. 2003.

[23] C. Nikou, G. Bueno, F. Heitz, and J. Armspach, "A joint physics-based statistical deformable model for multimodal brain image analysis," *IEEE Trans. Med. Imag.*, vol. 20, no. 10, pp. 1026–1037, Oct. 2001.

[24] M. Krinidis, N. Nikolaidis, and I. Pitas, "The discrete modal transform and its application to lossy image compression," *Signal Process.: Image Commun.*, vol. 22, no. 5, pp. 480–504, Jun. 2007.

[25] 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.

[26] B. Moghaddam, C. Nastar, and A. Pentland, "A bayesian similarity measure for direct image matching," in *Proc. Int. Conf. Pattern Recognition*, Vienna, Austria, Aug. 1996, pp. 350–358.

[27] A. Pentland and B. Horowitz, "Recovery of non-rigid motion and structure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 7, pp. 730–742, Jul. 1991.

[28] A. Pentland and B. Horowitz, "Recovery of non-rigid motion and structure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 7, pp. 730–742, Jul. 1991.

[29] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *IEEE Computers*, vol. C-23, pp. 90–93, Jan. 1974.

[30] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[31] Y. Deng, C. Kenney, M. S. Moore, and B. S. Manjunath, "Peer group filtering and perceptual color image quantization," in *Proc. IEEE Int. Symp. Circuits and Systems VLSI*, Orlando, FL, Jun. 1999, vol. 4, pp. 21–24.

[32] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression.*. Norwell, MA: Kluwer, 1992.

[33] R. O. Duda and R. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1970.

[34] C. Pantofaru and M. Hebert, A Comparison of Image Segmentation Algorithms Robotics Institute, Carnegie Mellon Univ., Tech. Rep. CMU-RI-TR-05-40, Sep. 2005.

[35] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Towards objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 929–944, Jun. 2007.

[36] M. Meila, "Comparing clusterings: An axiomatic view," in *Proc. Int. Conf. Machine Learning*, Bonn, Germany, Aug. 2005, pp. 577–584.

[37] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. Fifth Berkeley Symp. Mathematical Statistics and Probability*, Berkeley, CA, 1967, vol. 1, pp. 281–297.
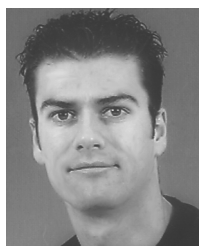
Michail Krinidis received the B.S. degree from the Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2002, where he is currently pursuing the Ph.D. degree while also serving as a teaching assistant.

His current research interests lie in the areas of 2-D tracking, face detection, and 3-D head pose estimation in image sequences.

Ioannis Pitas (S'83–M'84–SM'94–F'07) received the Diploma of Electrical Engineering in 1980 and the Ph.D. degree in electrical engineering in 1985 both from the Aristotle University of Thessaloniki, Greece.

Since 1994, he has been a Professor in the Department of Informatics, Aristotle University of Thessaloniki. From 1980 to 1993, he served as a Scientific Assistant, Lecturer, Assistant Professor, and Associate Professor in the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki. He has served as a Visiting Research Associate or Visiting Assistant Professor at several universities. He has published 153 journal papers, 400 conference papers, and contributed to 22 books in his areas of interest and edited or coauthored another five books. He has also been an invited speaker and/or member of the program committee of several scientific conferences and workshops.

Dr. Pitas has served as Associate Editor or Co-Editor of four international journals and as General or Technical Chair of three international conferences. His current interests are in the areas of digital image and video processing and analysis, multidimensional signal processing, watermarking, and computer vision.