Image Segmentation with a Parametric Deformable Model using Shape and Appearance Priors

Ayman El-Baz
Bioimaging Lab., Dept. of Bioengineering,
University of Louisville,
Louisville, KY, 40292, USA.
aselba01@louisville.edu

Georgy Gimel'farb
Dept. of Computer Science, Tamaki Campus,
The University of Auckland,
Auckland, New Zealand.
g.gimelfarb@auckland.ac.nz

Abstract

We propose a novel parametric deformable model controlled by shape and visual appearance priors learned from a training subset of co-aligned images of goal objects. The shape prior is derived from a linear combination of vectors of distances between the training boundaries and their common centroid. The appearance prior considers gray levels within each training boundary as a sample of a Markov-Gibbs random field with pairwise interaction. Spatially homogeneous interaction geometry and Gibbs potentials are analytically estimated from the training data. To accurately separate a goal object from an arbitrary background, empirical marginal gray level distributions inside and outside of the boundary are modeled with adaptive linear combinations of discrete Gaussians (LCDG). The evolution of the parametric deformable model is based on solving an Eikonal partial differential equation with a new speed function which combines the prior shape, prior appearance, and current appearance models. Due to the analytical shape and appearance priors and a simple Expectation-Maximization procedure for getting the object and background LCDG, our segmentation is considerably faster than most of the known geometric and parametric models. Experiments with various goal images confirm the robustness, accuracy, and speed of our approach.

1. Introduction

Parametric and geometric deformable models are widely used for image segmentation. However, in many applications, especially in medical image analysis, accurate segmentation with these models is a challenging problem due to noisy or low-contrast 2D/3D images with fuzzy boundaries between goal objects (e.g. anatomical structures) and their background; the similarly shaped objects with different visual appearances, and discontinuous boundaries because of occlusions or similar visual appearance of adjacent parts of objects of different shapes [1, 2]. Prior knowledge about the goal shape and/or visual appearance helps in solving such segmentation problems [2].

Relationship to the prior works: Conventional parametric deformable models [3] and geometric models (e.g. [4]) based on level set techniques [5] search for strong signal discontinuities (grayscale or color edges) in an image and do not account for prior shape constraints. But the evolution guided only by edges and general continuity–curvature limitations fails when a goal object is not clearly distinguishable from background. More accurate results, but at the expense of a considerably reduced speed, were obtained by restricting grayscale or color patterns within an evolving surface [6]. Nonetheless, the accuracy remains poor without prior knowledge of goal objects. At present, 2D/3D parametric and geometric deformable models with shape and/or appearance priors learned from a training set of manually segmented images are of the main interest.

Initial attempts to involve the prior shape knowledge were built upon the edges. Pentland and Sclaroff [7] and Cootes et al. [8] described an evolving curve with shape and pose parameters of a parametric set of points matched to strong image gradients and use a linear combination of eigenvectors to represent variations from an average shape. A parametric point model of Staib and Duncan [9] was based on an elliptic Fourier decomposition of landmarks. Model parameters ensure the best match between the evolving curve and points of strong gradients. Chakraborty et al. [10] extended this approach to a hybrid model combining the region gradient and homogeneity information.

More efficient results were obtained by learning the priors from a training set of manually segmented images of goal objects [11, 12, 13, 14, 1]. Pizer et al. [11] and Styner et al. [13] segment 3D medical images by coarse-to-fine deformation of a shape-based medial representation ("m-rep"). A deformable model of Huang et al. [15] integrates region, shape and interior signal features assuming an ap-

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

proximate region shape is a priori known and aligned with the image to initialize the model. Leventon et al. [12] and Shen and Davatzikos [14] augment a level set-based energy function guiding the evolution with special terms attracting to more likely shapes specified with the principal component analysis (PCA) of the training set of goal objects, while Chen et al. [16] use a geometric model with the prior "average shape". The most advanced level set-based geometric model of Tsai et al. [1] evolves as zero level of a 2D map of the signed shortest distances between each pixel and the boundary. The goal shapes are approximated with a linear combination of the training distance maps for a set of mutually aligned training images. High dimensionality of the distance map space hinders PCA, and to simplify the model, only a few top-rank principal components are included to the linear combination.

Prior appearance models have not been explicitly used to guide the evolution of all above-mentioned models. Typically, the guidance was based on a simple predefined appearance model assuming significantly different gray level means or variances in an object and background (see e.g. [1]). More promising is to learn the appearance model as well, although the segmentation in this case involves usually only pixel- or voxel-wise classification [17]. An alternative approach of Joshi [18] performs nonparametric warping of a goal surface to a deformable atlas. The atlas contours are then transferred to the goal volume. But due to no shape prior, the resulting boundaries need not approach the actual ones. Speed of the iterative warpings of a whole image volume is also quite low.

To overcome these problems, Paragios and Deriche [19] and Cootes et al. [20] learned the joint probabilistic shape–appearance prior models of goal objects with the PCA. A few simpler such models have been successfully applied to segment complex 3D medical images [21]. But segmentation still was slow due to setting up pixel- or voxel-wise model–image correspondences at each step. To accelerate the process, Yang and Duncan [22] and Freedman et al. [23] introduced probabilistic appearance priors accounting for only low-order signal statistics. Yang and Duncan [22] used a joint Gaussian probability model to describe variability of the goal shapes and image gray levels and developed optimization algorithms for estimating model parameters from a set of training images and conducting the maximum a posteriori (MAP) segmentation. Freedman et al. [23] derived a special level-set function such that its zero level produces approximately the goal shapes and use empirical marginal gray level distributions for the training objects as the appearance prior to guide the alignment of a deformable model to an image to be segmented. Foulonneau et al. [24] presented a new algorithm to constrain the evolution of a region-based active contour with respect to a reference shape. Their shape prior model includes intrinsic invariance with regard to pose and affine deformations.

Our approach follows the same ideas of using both the shape and appearance prior knowledge, but differs in the three aspects. **First**, instead of using the level set framework running into problems with linear combinations of the distance maps, we use a simple parametric deformable model. Both the model and each goal shape are represented by piecewise-linear boundaries with a predefined number of control points. Corresponding points are positioned on roughly equiangular rays from the common center being the centroid of the control points along each boundary. A robust wave propagation is used to find correspondences in an aligned pair of the boundaries. **Secondly**, visual appearance of the goal objects is roughly described by characteristic statistics of gray level co–occurrences. Grayscale object pattern is considered as a sample of a spatially homogeneous Markov-Gibbs random field (MGRF) with multiple pairwise interaction. The interaction parameters of MGRF are estimated analytically. **Third**, the evolution is also guided at each step with a first-order probability model of the current appearance of a goal object and its background.

Basic notation:

- $\mathbf{R} = [(x, y) : x = 0, \ldots, X-1; y = 0, \ldots, Y-1]$ – a finite arithmetic lattice supporting digital images and their region maps.
- $\mathbf{g} = [g_{x,y} : (x, y) \in \mathbf{R}; g_{x,y} \in \mathcal{Q}]$ – a grayscale image taking values from a finite set $\mathcal{Q} = \{0, ..., Q-1\}$.
- $\mathbf{m} = [m_{x,y} : (x, y) \in \mathbf{R}; m_{x,y} \in \mathbf{L}]$ – a region map taking labels from a binary set $\mathbf{L} = \{\mathsf{ob}, \mathsf{bg}\}$; each label $m_{x,y}$ indicates whether the pixel (x, y) in the corresponding image $\mathbf{g}$ belongs to a goal object (ob) or background (bg).
- $\mathbf{b} = [\mathbf{p}_k : k = 1, \ldots, K]$ – a deformable piecewise-linear boundary with the K control points $\mathbf{p}_k = (x_k, y_k)$ forming a circularly connected chain of line segments $(\mathbf{p}_1, \mathbf{p}_2), \ldots, (\mathbf{p}_{K-1}, \mathbf{p}_K), (\mathbf{p}_K, \mathbf{p}_1)$.
- $\mathbf{d} = [d_k^2 : k = 1, \ldots, K]$ – a vector description of the boundary $\mathbf{b}$ in terms of the square distances $d_k^2 = (x_k - x_0)^2 + (y_k - y_0)^2$ from the control points to the model centroid $\mathbf{p}_0 = (x_0 = \frac{1}{K}\sum_{k=1}^{K} x_k, y_0 = \frac{1}{K}\sum_{k=1}^{K} y_k)$, i.e. to the point at the minimum mean square distance from all the control points.
- $\mathbf{S} = \{(\mathbf{g}_t, \mathbf{m}_t, \mathbf{b}_t, \mathbf{d}_t) : t = 1, \ldots, T\}$ – a training set of grayscale images of the goal objects with manually prepared region maps and boundary models.
- $|\mathbf{A}|$ – the cardinality of a finite set $\mathbf{A}$.

2. Shape Prior

To build the shape prior, all the training objects in $\mathbf{S}$ are mutually aligned to have the same centroid and unified poses (orientations and scales of the objects boundaries) as in Fig. 1(a). For the definiteness, let each training boundary $\mathbf{b}_t \in \mathbf{S}$ is represented with K control points on the polar system of K° equiangular rays (i.e. with the angular pitch $2\pi/K^\circ$) emitted from the common centroid $\mathbf{p}_0$. The rays are enumerated clockwise, with zero angle for the first position $\mathbf{p}_{t,1}$ of each boundary. Generally, there may be rays with no or more than one intersection of a particular boundary, so that the number of the control points K may differ from the number of the rays K°.

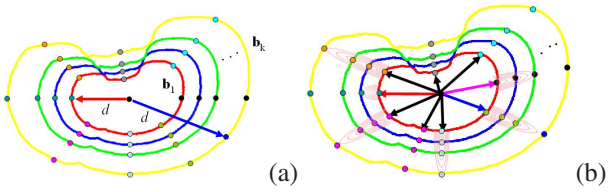


Figure 1. (a) Mutually aligned training boundaries and (b) searching for corresponding points between the mutually aligned training boundaries.

Because the training boundaries $\mathbf{b}_t \in \mathbf{S}$; $t = 1, \ldots, T$, share the same centroid $\mathbf{p}_0$, any linear combination $\mathbf{d} = \sum_{t=1}^{T} w_t \mathbf{d}_t$ of the training distance vectors defines a unique new boundary $\mathbf{b}$ with the same centroid. Typically, shapes of the training objects are very similar, and their linear combinations could be simplified by the PCA to escape singularities when adjusting to a given boundary.

Let $\mathbf{D} = [\mathbf{d}_1 \ \mathbf{d}_2 \ \cdots \mathbf{d}_T]$ and $\mathbf{U} = \mathbf{D}\mathbf{D}^\mathsf{T}$ denote the $K \times T$ matrix with the training distance vectors as columns and the symmetric $K \times K$ Gram matrix of sums of squares and pair products $\sum_{t=1}^{T} d_{t,k} d_{t,k'}$; $k, k' = 1, \ldots, K$ of their components, respectively. The PCA of the matrix $\mathbf{U}$ produces K eigen-vectors $[\mathbf{e}_i : i = 1, \ldots, K]$ sorted by their eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K \geq 0$. Due to identical or very similar training shapes, most of the bottom-rank eigenvalues are zero or very small, so that the corresponding "noise" eigenvectors can be discarded. Only a few top-rank eigenvectors actually represent the training shapes; the top distance eigenvector $\mathbf{e}_1$ corresponds to an "average" shape and a few others determine its basic variability. For simplicity, we select the top-rank subset of the eigenvectors $(\mathbf{e}_i : i = 1, \ldots, K')$; $K' < K$ by thresholding: $\sum_{i=1}^{K'} \lambda_i \approx \theta \sum_{i=1}^{K} \lambda_i$ with an empirical threshold $\theta = 0.8 \ldots 0.9$.

An arbitrary boundary $\mathbf{b}_\mathrm{c}$ aligned with the training set is described with the vector $\mathbf{d}_\mathrm{c}$ of the squared distances from its control points to the centroid. The prior shape approximating this boundary is specified by the linear combination of the training vectors: $\mathbf{d}^* = \sum_{i=1}^{K'} w_i^* \mathbf{e}_i \equiv \sum_{i=1}^{K'} (\mathbf{e}_i^\mathsf{T} \mathbf{d}_\mathrm{c}) \mathbf{e}_i$. Each signed difference $\Delta_k = d_k^* - d_{\mathrm{c},k}$ determines the direction and force to move the boundary $\mathbf{b}_\mathrm{c}$ towards the closest shape prior $\mathbf{b}^*$ specified by the distance vector $\mathbf{d}^*$.

Search for corresponding points is performed to suppress local "noise" (spurious deviations) in the training boundaries. The corresponding points are found in the aligned training boundaries by a robust wave-propagation based search (see Fig. 1(b)). An orthogonal wave is emitted from a point in one boundary, and the point at which the maximum curvature position of the wave front hits the second boundary is considered as the corresponding point (see Fig. 2). The whole search algorithm is as follows:

1. Given a set of aligned training boundaries, pick any one as the reference shape, A.
2. Represent the reference shape with K equiangular points.
3. Find the corresponding point in each other training boundary, B, for every point representing the reference shape A:
   (a) Find the intersection area between the shapes B and A.
   (b) Find the normalized minimum Euclidian distance $D(x, y)$ between every point (x, y) in the intersection area and the boundary A by solving the Eikonal equation:
   $$|\nabla T(x, y)| F(x, y) = 1 \qquad (1)$$
   where $T(x, y)$ is the time at which the front crosses the point (x, y); the solution uses the fast marching level sets at unity speed function, $F(x, y) = 1$ [25].
   (c) For every point representing the reference shape A, generate an orthogonal wave by solving the Eikonal equation using the fast marching level sets at the speed function $F(x, y) = \exp(D(x, y))$. For more detail see [26].
   (d) Track the point with the maximum curvature for each propagating wave front (see Figs. 2); this point is considered at any time as the corresponding point to the starting one on the reference shape A.
   (e) The point at which the maximum curvature point of the propagating wave hits the boundary B is selected as the point corresponding to the starting point of the reference shape A.

SIFT-based alignment: Just as the conventional level-set based geometric models with the shape priors, e.g. in [1], our approach depends on accuracy of mutual alignment of similar shapes at both the training and segmentation stages.

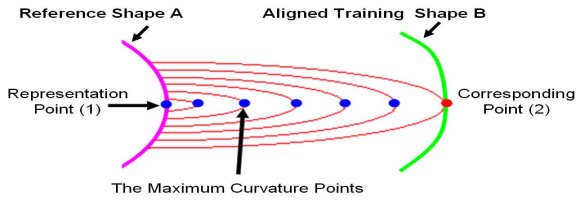
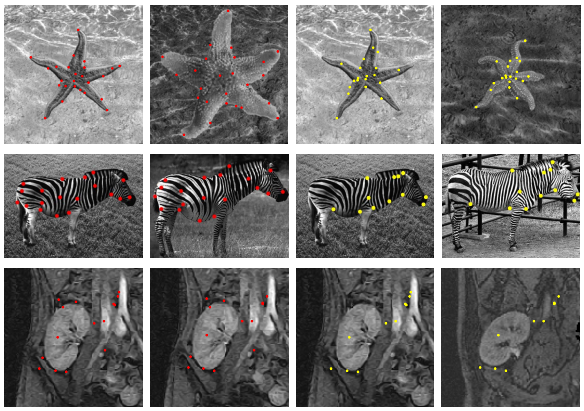


Figure 2. The point 2 at which the maximum curvature point of the wave hits the boundary B corresponds to the starting point 1 on the reference shape A.



(a–b) (a–c)

Figure 3. Corresponding points found by SIFT in each pair (a,b), and (a,c) shown in red and yellow, respectively.

In the latter case the deformable model is initialized by aligning an image $\mathbf{g}$ to be segmented with one of the training images, say, $\mathbf{g}_1 \in \mathbf{S}$, arbitrarily chosen as a prototype.

First we use the *scale invariant feature transform* (SIFT) proposed by Lowe [27] to reliably determine a number of point-wise correspondences between two images under their relative affine geometric and local contrast / offset signal distortions. Then the affine transform aligning $\mathbf{g}$ most closely to $\mathbf{g}_1$ is determined by the gradient descent minimization of the mean squared positional error between the corresponding points.

Figure 3 shows correspondences found by SIFT in digital natural images of starfish, zebras and magnetic resonance images (MRI) of human kidneys. The resulting affinely aligned goal shapes have roughly the same center and similar poses (orientations and scales). Quality of such alignment is evaluated in Fig. 4 by averaging all the training region maps $\mathbf{m}_t$; $t = 1, \ldots, T$, before and after the training set $\mathbf{S}$ is mutually aligned.

3. Appearance Models

MGRF-based appearance prior. Our appearance prior is a rough descriptor of typically complex grayscale patterns of goal objects in terms of only second-order signal statistics. Each goal image is considered as a sample of a pairwise Markov–Gibbs random field (MGRF).

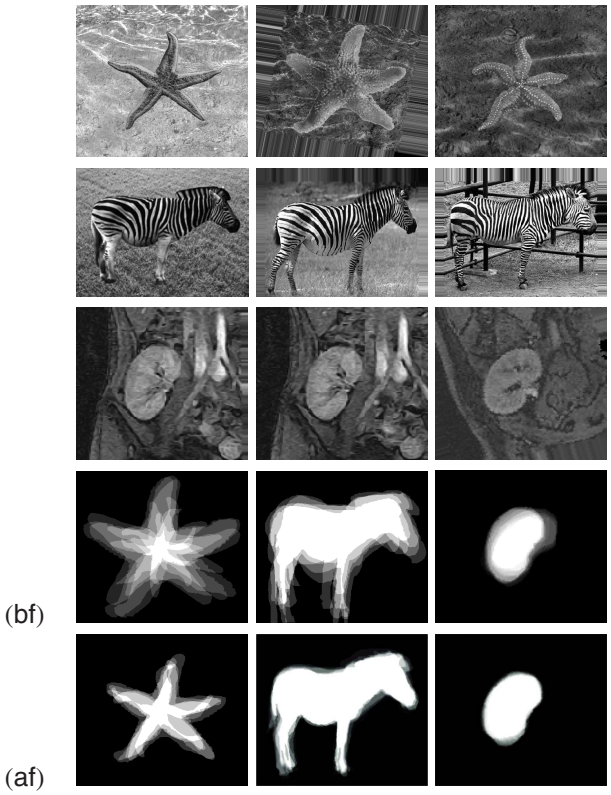


Figure 4. Mutually aligned images shown in Fig. 3 (three upper rows), and overlaps of the training region maps before (bf) and after (af) the alignment.

Let $\mathcal{N} = \{(\xi_i, \eta_i) : i = 1, \ldots, n\}$ be a finite set of (x, y)-offsets specifying neighbors $\{((x+\xi, y+\eta), (x-\xi, y-\eta)) : (\xi, \eta) \in \mathcal{N}\}$ $\wedge$ $\mathbf{R}$ interacting with each pixel $(x, y) \in \mathbf{R}$. Let $\mathcal{C}_{\xi,\eta}$ be a family of pairs $c_{\xi,\eta;x,y} = ((x, y), (x+\xi, y+\eta))$ in $\mathbf{R}$ with the offset $(\xi, \eta) \in \mathcal{N}$, i.e. the family of translation invariant pairwise cliques of the neighborhood graph on $\mathbf{R}$. Let $\mathbf{V}$ be a vector of Gibbs potentials for gray level co-occurrences in the neighboring pairs: $\mathbf{V}^{\mathsf{T}} = [\mathbf{V}^{\mathsf{T}}_{\xi,\eta} : (\xi, \eta) \in \mathcal{N}]$ where $\mathbf{V}^{\mathsf{T}}_{\xi,\eta} = [V_{\xi,\eta}(q, q') : (q, q') \in \mathcal{Q}^2]$.

A generic MGRF with multiple pairwise interaction on $\mathbf{R}$ is specified by the Gibbs probability distribution (GPD)

$$P(\mathbf{g}) = \frac{1}{Z} \exp \sum_{(\xi,\eta)\in\mathcal{N}} \sum_{c_{\xi,\eta;x,y}\in\mathcal{C}_{\xi,\eta}} V_{\xi,\eta}(g_{x,y}, g_{x+\xi,y+\eta}) \equiv \frac{1}{Z} \exp |\mathbf{R}| \mathbf{V}^{\mathsf{T}} \mathbf{F}(\mathbf{g}) \quad (2)$$

Here, Z is the partition function, $\mathbf{F}^{\mathsf{T}}(\mathbf{g})$ is the vector of scaled empirical probability distributions of gray level co-occurrences over each clique family: $\mathbf{F}^{\mathsf{T}}(\mathbf{g}) = [\rho_{\xi,\eta}\mathbf{F}^{\mathsf{T}}_{\xi,\eta}(\mathbf{g}) : (\xi, \eta) \in \mathcal{N}]$ where $\mathbf{F}_{\xi,\eta}(\mathbf{g}) = [f_{\xi,\eta}(q, q'|\mathbf{g}) : (q, q') \in \mathcal{Q}^2]^{\mathsf{T}}$ and $\rho_{\xi,\eta} = \frac{|\mathcal{C}_{\xi,\eta}|}{|\mathbf{R}|}$ is the relative size of the clique family.

The empirical probabilities are $f_{\xi,\eta}(q, q'|\mathbf{g}) =$

$\frac{|\mathcal{C}_{\xi,\eta;q,q'}(\mathbf{g})|}{|\mathcal{C}_{\xi,\eta}|}$ where $\mathcal{C}_{\xi,\eta;q,q'}(\mathbf{g})$ is the subfamily of the pairs $c_{\xi,\eta;x,y} \in \mathcal{C}_{\xi,\eta}$ supporting a co-occurrence $(g(x,y)=q, g(x+\xi,y+\eta)=q')$ in the image $\mathbf{g}$.

To specify the appearance prior, let $\mathbf{R}_t = \{(x,y) : (x,y) \in \mathbf{R} \wedge m_{t;x,y} = \mathsf{ob}\}$ be a part of $\mathbf{R}$ supporting the goal object in the training image–map pair $(\mathbf{g}_t, \mathbf{m}_t) \in \mathbf{S}$. Let $\mathcal{C}_{\xi,\eta;t} \subset \mathcal{C}_{\xi,\eta}$ and $\mathbf{F}_{\xi,\eta;t}$ denote the subfamily of the pixel pairs in $\mathbf{R}_t$ with the coordinate offset $(\xi,\eta) \in \mathcal{N}$ and the empirical probability distribution of gray level co-occurrences in the training image $\mathbf{g}_t$ over the subfamily $\mathcal{C}_{\xi,\eta;t}$, respectively. The model of the t-th training object has the GPD on the sublattice $\mathbf{R}_t$:

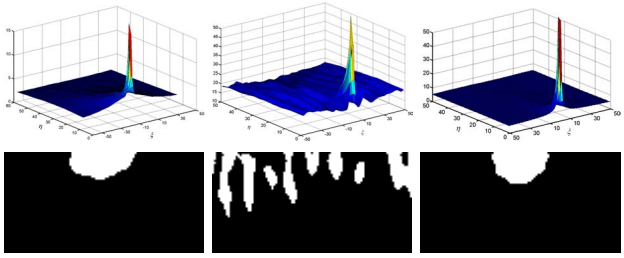
$$P_t = \frac{1}{Z_t}\exp\left(|\mathbf{R}_t| \sum_{(\xi,\eta)\in\mathcal{N}} \rho_{\xi,\eta;t}\mathbf{V}^{\mathsf{T}}_{\xi,\eta}\mathbf{F}_{\xi,\eta;t}\right) \quad (3)$$

where $\rho_{\xi,\eta;t} = |\mathbf{C}_{\xi,\eta;t}|/|\mathbf{R}_t|$ is the relative size of $\mathcal{C}_{\xi,\eta;t}$ with respect to $\mathbf{R}_t$.

The areas and shapes of the aligned goal objects are similar for all $t = 1,\ldots,T$, so that $|\mathbf{R}_t| \approx R_{\mathsf{ob}}$ and $|\mathcal{C}_{\xi,\eta;t}| \approx C_{\mathsf{ob};\xi,\eta}$ where R_{ob} and $C_{\nu,\mathsf{ob}}$ are the average cardinalities over the training set $\mathbf{S}$. Assuming the independent samples, all the training objects are represented with the joint GPD:

$$P_{\mathbf{S}} \approx \frac{1}{Z_{\mathbf{S}}}\exp\left(TR_{\mathsf{ob}} \sum_{(\xi,\eta)\in\mathcal{N}} \rho_{\mathsf{ob};\xi,\eta}\mathbf{V}^{\mathsf{T}}_{\mathsf{ob};\xi,\eta}\mathbf{F}_{\mathsf{ob};\xi,\eta}\right) \quad (4)$$

where $\rho_{\mathsf{ob};\xi,\eta} = \mathcal{C}_{\mathsf{ob};\xi,\eta}/R_{\mathsf{ob}}$, and the empirical probability distributions $\mathbf{F}_{\mathsf{ob};\xi,\eta}$ describe the gray level co-occurrences in all the training goal objects.



(a) $|\mathcal{N}'| = 61$ (b) $|\mathcal{N}'| = 168$ (c) $|\mathcal{N}'| = 76$

Figure 5. Relative interaction energies for the clique families in function of the offsets (ξ,η) and the characteristic pixel neighbors $\mathcal{N}'$ (white areas in the (ξ,η)-plane) for the training sets of 43 starfish (a), 67 zebra (b), and 1300 kidney (c) images.

We use here our novel analytical maximum likelihood estimator for the Gibbs potentials of the pair-wise MGRF [28]:

$$V_{\mathsf{ob};\xi,\eta}(q,q') = \lambda\rho_{\mathsf{ob};\xi,\eta}[f_{\mathsf{ob};\xi,\eta}(q,q') - f_{\mathsf{ob}}(q)f_{\mathsf{ob}}(q')] \quad (5)$$

where $f_{\mathsf{ob}}(.)$ and $f_{\mathsf{ob};\xi,\eta}(q,q')$ be a joint empirical probability distribution of pixel intensities and of intensity co-occurrences, respectively. λ is the analytically computed common scaling factor: $\lambda \approx Q^2$ if $Q \gg 1$, $f_{\mathsf{ob}}(q) \approx 1/Q$ and $\rho_{\mathsf{ob};\xi,\eta} \approx 1$ for all $(\xi,\eta) \in \mathcal{N}$. It can be omitted, i.e., set to $\lambda = 1$, when only relative potential values are of interest, e.g. to rank relative Gibbs energies $E^{\mathrm{rel}}_{\mathsf{ob};\xi,\eta}$ of pairwise interaction in the goal objects, i.e. the variances of the co-occurrence distributions: $E^{\mathrm{rel}}_{\mathsf{ob};\xi,\eta} = \sum_{q,q'\in\mathcal{Q}^2} f_{\mathsf{ob};\xi,\eta}(q,q')[f_{\mathsf{ob};\xi,\eta}(q,q') - f_{\mathsf{ob}}(q)f_{\mathsf{ob}}(q')]$. Most characteristic neighbors $\mathcal{N}' \subset \mathcal{N}$ are selected as the prior appearance descriptors in Eq. (5) by thresholding the relative energies [29]. Figure 5 shows distributions of the relative energies and the selected neighbors.

Under this prior description, a grayscale pattern within each current deformable boundary $\mathbf{b}$ in an image $\mathbf{g}$ is described by its relative Gibbs energy $E(\mathbf{g},\mathbf{b}) = \sum_{(\xi,\eta)\in\mathcal{N}'} \mathbf{V}^{\mathsf{T}}_{\mathsf{ob};\xi,\eta}\mathbf{F}_{\xi,\eta}(\mathbf{g},\mathbf{b})$ where $\mathcal{N}'$ is a selected subset of the top-rank neighbors, and the empirical distributions $\mathbf{F}_{\xi,\eta}(\mathbf{g},\mathbf{b})$ are collected in $\mathbf{b}$.

LCDG-models of current appearance: To more accurately account for the current image appearance in addition to the learned shape and appearance priors, 1D empirical marginal gray level distributions inside and outside of an initial deformable boundary $\mathbf{b}$ are approximated with linear combinations of discrete Gaussians (LCDG). A DG $\mathbf{\Psi}_\theta = (\psi(q|\theta) : q \in \mathcal{Q})$ is defined as a discrete probability distribution with components integrating a continuous Gaussian density over intervals related to successive gray levels in $\mathcal{Q}$, that is, $\psi(0|\theta) = \Phi_\theta(0.5)$, $\psi(q|\theta) = \Phi_\theta(q+0.5) - \Phi_\theta(q-0.5)$ for $q = 1,\ldots,Q-2$, and $\psi(Q-1|\theta) = 1 - \Phi_\theta(Q-1.5)$ where $\Phi_\theta(q)$ is the cumulative Gaussian probability function with a shorthand notation $\theta = (\mu,\sigma^2)$ for its mean, μ, and variance, σ^2.

The numbers $K_l; l \in \mathbf{L}$, of dominant DGs in each model are determined by maximizing the Akaike Information Criterion (AIC) of the corresponding empirical distributions. Then the LCDG-models with positive dominant and sign-alternate subordinate DGs are built with the EM-based techniques introduced in [30]. The subordinate positive and negative DGs approximate deviations of the empirical distribution from the dominant mixture. Let K_{p}; $K_{\mathrm{p}} \geq K_l$, and K_{n} denote the total numbers of the positive and negative components. Then the LCDG model is as follows:

$$p_{\mathbf{w},\mathbf{\Theta}}(q) = \sum_{r=1}^{K_{\mathrm{p}}} w_{\mathrm{p},r}\psi(q|\theta_{\mathrm{p},r}) - \sum_{l=1}^{K_{\mathrm{n}}} w_{\mathrm{n},l}\psi(q|\theta_{\mathrm{n},l}) \quad (6)$$

where the non-negative weights $\mathbf{w} = [w_{p,.}, w_{n,.}]$ meet the obvious restriction $\sum_{r=1}^{K_{\mathrm{p}}} w_{\mathrm{p},r} - \sum_{l=1}^{K_{\mathrm{n}}} w_{\mathrm{n},l} = 1$. To identify this model, the numbers $K_{\mathrm{p}} - K_l$ and K_{n} of its subordinate components and the parameters $\mathbf{w}$, $\mathbf{\Theta}$ (weights, means, and variances) of all the DGs are estimated first with a sequential EM-based initializing algorithm producing a close initial LCDG-approximation of the empirical

distribution. Then under the fixed numbers $K_{\rm p}$ and $K_{\rm n}$, all other parameters are refined with a modified EM algorithm that accounts for the alternating components. Each final LCDG-model is partitioned into two LCDG-submodels $\mathbf{p}_{{\rm pix},l} = [p_{{\rm pix},l}(q) : q \in \mathcal{Q}]$, one per class $l \in \mathbf{L}$, by associating the subordinate DGs with the corresponding dominant terms such that the misclassification rate is minimal.

4. Model Evolution

The evolution $\mathbf{b}_\tau \rightarrow \mathbf{b}_{\tau+1}$ of the deformable boundary $\mathbf{b}$ in discrete time, $\tau = 0, 1, \ldots$, is determined by solving the Eikonal equation $|\nabla T(\mathbf{p}_{k,\tau})|F(\mathbf{p}_{k,\tau})$; $k = 1, \ldots, K$, where $F(\mathbf{p})$ is a speed function for the control point $\mathbf{p} = [x, y]$ of the current boundary. Our speed function depends on the shape prior, the LCDG-model of current appearance, and the MGRF-based appearance prior:

$$F(\mathbf{p}) = e^{-\beta|\Delta|} p_{{\rm pix,ob}}(g_{x,y}) \pi_{x,y}(g_{x,y}|\mathbf{S}) \qquad (7)$$

Here, Δ is the signed distance between the current control point $\mathbf{p} \in \mathbf{b}_\tau$ and the like one in the closest shape prior along the ray from the current boundary centroid. Note that the sign of Δ determines the direction of propagation as shown in Fig. 6. The constant factor β determines the evolution speed ($0 < \beta < 1$ for a smooth propagation). The marginal probabilities $p_{{\rm pix,ob}}(q)$ and $p_{{\rm pix,bg}}(q)$ of the gray value $g_{x,y} = q$ are estimated with the LCDG-submodels for the object and its background, respectively. The prior conditional probability $\pi_{x,y}(q|\mathbf{S})$ of the gray value $g_{x,y} = q$ in the pixel $\mathbf{p} = (x, y)$, given the current gray values in its neighborhood, is estimated in line with the MGRF prior appearance model:

$$\pi_{x,y}(g_{x,y}|\mathbf{S}) = \frac{\exp\left(E_{x,y}(g_{x,y}|\mathbf{S})\right)}{\sum\limits_{q \in \mathcal{Q}} \exp\left(E_{x,y}(q|\mathbf{S})\right)} \qquad (8)$$

where $E_{x,y}(q|\mathbf{S})$ is the pixel-wise Gibbs energy for the gray value q in the pixel $\mathbf{p} = (x, y)$, given the fixed gray values in its characteristic neighborhood: $E_{x,y}(q|\mathbf{S}) =$

$$\sum_{(\xi,\eta) \in \mathbf{N}'} \left(V_{{\rm ob};\xi,\eta}(g_{x-\xi,y-\eta}, q) + V_{{\rm ob};\xi,\eta}(q, g_{x+\xi,y+\eta})\right)$$

In total, the proposed segmentation algorithm is as follows:

1. Initialization ($\tau = 0$):
   (a) Find the affine alignment of a given image $\mathbf{g}$ to a selected prototype $\mathbf{g}_1 \in \mathbf{S}$ using the SIFT correspondences and gradient optimization.
   (b) Initialize the deformable model with the training boundary $\mathbf{b}_1$ for $\mathbf{g}_1$.
   (c) Find the current appearance LCDG model.
2. Evolution ($\tau \leftarrow \tau + 1$):
   (a) Evolve the parametric deformable model boundary based on solving the Eikonal PDE with the speed function shown in Eq. (7).
   (b) Terminate the process if the overall absolute deformation $\sum_{k=1}^{K} |d_{k,\tau+1} - d_{k,\tau}| \leq \alpha$ (a small predefined threshold); otherwise return to Step 2a.
3. Segmentation: transfer the final boundary to the initial (non-aligned) image $\mathbf{g}$ by the inverse affine transform.

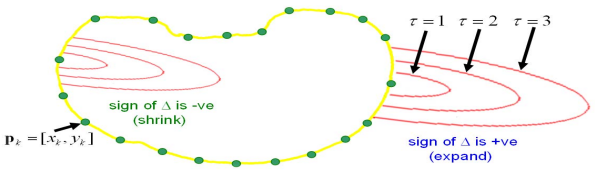


Figure 6. Illustration of the propagation of a parametric deformable model using fast marching level set [25].

5. Experimental Results and Conclusions

The performance of the proposed parametric deformable model was evaluated on a large number of intricate digital images such as starfish and zebras with the visually obvious ground truth (actual object boundaries) and dynamic contrast-enhanced MRI (DCE-MRI) of human kidneys with the ground truth presented by a radiologist. The DCE-MR images are usually noisy, with continuously changing and low contrast. About 33% of images of each type were used to learn the priors.

Basic segmentation stages of the algorithm are shown in Fig. 7. To show the effect of each model (shape, learned prior appearance model, and current appearance model) in the final segmentation results, we calculated the pixel-wise energy for each pixel in the image as shown in Figs. 7(e, f, g, h, i). It is clear from Figs. 7(e, f, g, h, i) the pixel-wise energies for the object (starfish) are higher than any other pixels appearing in the background, which confirms good guidance for the evolution of the deformable model. Figure 7(e) shows the pixel-wise Gibbs energy for each pixel in the given image. The segmentation based only on the learned prior appearance model (Gibbs energy) is shown in Fig 7(f). The major drawback of using only Gibbs energy as a guidance for the deformable model evolution is the error in the segmentation which occurs at the edges of the learned object due to the Gibbs interaction model. The current appearance model is used to reduce the blurring effect at the edges and increase the contrast as shown in Figure 7(g). It is clear from Fig. 7(h) the segmentation based on learned prior and current appearance models still contains errors around the edges of the starfish object, which can be minimized by adding the shape prior model to these two pervious models as shown in Fig. 7(i, j). Figures 8 and 9 highlight the

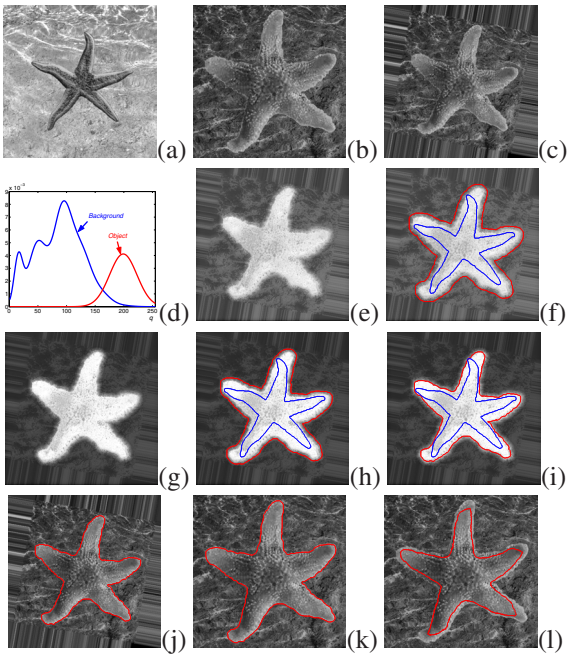


Figure 7. Chosen training prototype (a); an image to be segmented (b); its alignment to the prototype (c); LCDG-estimates (d) of the marginal distributions $\mathbf{p}_{\text{pix,ob}}$ and $\mathbf{p}_{\text{pix,bg}}$; pixel-wise Gibbs energies (learned prior appearance model)(e), segmentation result using learned prior appearance model (initialization shown in blue color and final segmentation shown in red color of the image aligned to the training prototype (f), pixel-wise Gibbs energy (learned prior appearance model) and LCDG energy (current appearance model) (g), segmentation results based on learned prior and current appearance models (h), segmentation results based on learned prior shape, and prior and current appearance models imposed in the energy domain (i) and in image domain (j). The final result (k) after its inverse affine transform to the initial image (b) (the total error 0.74% compared to the ground truth), and the final segmentation (h) with the algorithm in [1] (the total error 11.6%). *Note that our results obtained using 525 control points describing starfish shape*.

Figure 8. (a) an image to be segmented (note that the starfish is occluded by a leaf), (b) segmentation results based on learned prior and current appearance models imposed on the energy domain, and the final segmentation results based on learned prior shape, and prior and current appearance models imposed in the energy domain (c) and in the image domain (d) (the total error 1.27% compared to the ground truth). *Note that our results obtained using 525 control points describing starfish shape*.

advantages of using the learned prior shape model in addition to the learned visual appearance models in case the object is occluded by another object as shown in Fig. 8, or two overlapping objects having the same visual appearance

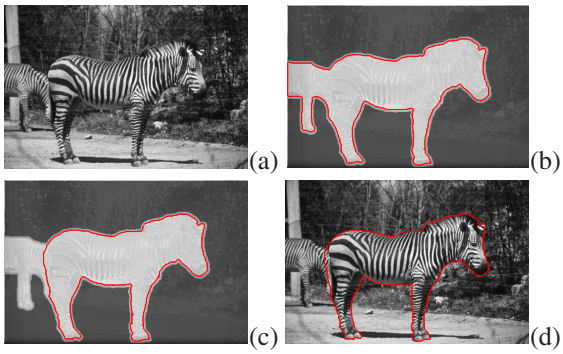


Figure 9. (a) an image to be segmented (note that the two zebras are overlapping and have the same visual appearance), (b) segmentation results based on learned prior and current appearance models imposed on the energy domain, and the final segmentation results based on learned prior shape, and prior and current appearance models imposed in the energy domain (c) and in image domain (d) (the total error 2.03% compared to the ground truth). *Note that our results obtained using 927 control points describing zebra shape*.

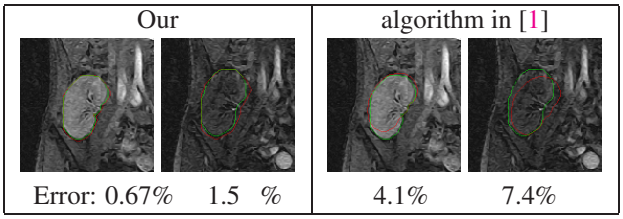


Figure 10. More segmentation of two kidney DCE-MR images with our approach and the algorithm in [1] vs. the ground truth: the final boundaries and the ground truth are in red and green, respectively. *Note that our results obtained using 140 control points describing kidney shape*.

models as shown in Fig. 9. Additional experimental results for kidney images are shown in Figure 10. Due to space limitations, more results for starfish, zebra and kidney images and comparisons with other approaches (e.g. active shape model, . . .) are provided on our website[1].

Experiments with different natural images provide support for the proposed parametric deformable model guided with the learned shape and appearance priors. Our approach assumes that (*i*) the boundaries of the training and test objects are reasonably similar to within a relative affine transform and (*ii*) SIFT reliably detects corresponding points to automatically align the goal objects in the images in spite of their different backgrounds. Although these assumptions restrict an application area of our approach comparing to the conventional parametric models, the latter typically fail on the above and similar images. More accurate level set-based geometric models with linear combinations of the training distance maps as the shape priors also rely on the mutual image alignment. Compared to these models, our approach escapes some of theoretical inconsistencies,

[1] http://uofl.edu/speed/bioengineering/faculty/bioengineering-full/dr-ayman-el-baz/elbazlab.html.

is computationally much simpler and faster, and has similar accuracy on high-contrast images, but notably better performance on low-contrast and multimodal ones.

References

[1] A. Tsai, A. Yezzi, J. Wells, C. Tempany, D. Tucker, A. Fan, W. Grimson, and A. Willsky, "A shape based approach to the segmentation of medical imagery using level sets," *IEEE Trans. Medical Imaging*, vol. 22, pp. 137–154, 2003.

[2] M. Rousson and N. Paragios, "Shape priors for level set representations," *Proc. 7$^{\text{th}}$ European Conf. Computer Vision*, Denmark, June 2002, pp. 78–92.

[3] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Computer Vision*, vol. 1, pp. 321–331, 1987.

[4] V. Caselles, R. Kimmel, and G. Sapiro, "On geodesic active contours," *Int. J. Computer Vision*, vol. 22, pp. 61–79, 1997.

[5] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulation," *J. Comput. Physics*, vol. 79, pp. 12–49, 1988.

[6] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, vol. 10, pp. 266–277, 2001.

[7] A. Pentland and S. Sclaroff, "Closed-form solutions for physically based shape modeling and recognition," *IEEE Trans. PAMI*, vol. 13, pp. 715–729, 1991.

[8] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision Image Understanding*, vol. 61, pp. 38–59, 1995.

[9] L. Staib and J. Duncan, "Boundary finding with parametrically deformable contour models," *IEEE Trans. PAMI*, vol. 14, pp. 1061–1075, 1992.

[10] A. Chakraborty, L. Staib, and J. Duncan, "An integrated approach to boundary finding in medical images," *Proc. IEEE Workshop Biomedical Image Analysis*, Seattle, Washington, June 24–25, 1994, pp. 13–22.

[11] S. Pizer, G. Gerig, S. Joshi, and S. Aylward, "Multiscale medial shape-based analysis of image objects," *Proc. IEEE*, vol. 91, pp. 1670–1679, 2003.

[12] M. Leventon, W. Eric, L. Grimson, O. Faugeras, "Statistical shape influence in geodesic active contours," *In Proc. CVPR*, South Carolina, June 13–15, 2000, pp. 316–323.

[13] M. Styner, G. Gerig, S. Pizer, and S. Joshi, "Automatic and robust computation of 3D medial models incorporating object variability," *Int. J. Computer Vision*, vol. 55, pp. 107–122, 2002.

[14] D. Shen and C. Davatzikos, "An adaptive-focus deformable model using statistical and geometric information," *IEEE Trans. PAMI*, vol. 22, pp. 906–913, 2000.

[15] X. Huang, D. Metaxas, and T. Chen, "MetaMorphs: Deformable shape and texture models," *In Proc. CVPR*, Washington, D.C., June 27 – July 2, 2004, vol. 1, pp. 496–503.

[16] Y. Chen, S. Thiruvenkadam, F. Huang, D. Wilson, E. Geiser, and H. Tagare, "On the incorporation of shape priors into geometric active contours," *Proc. IEEE Workshop Variational and Level Set Methods*, Canada, July 2001, pp. 145–152.

[17] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "A unifying framework for partial volume segmentation of brain MR images," *IEEE Trans. Medical Imaging*, vol. 22, no. 1, pp. 105–119, 2003.

[18] S. Joshi, "Large deformation diffeomorphisms and Gaussian random fields for statistical characterization of brain submanifolds," *PhD Thesis*, Washington Univ., 1997.

[19] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Trans. PAMI*, vol. 22, pp. 266–280, 2000.

[20] T. Cootes and C. Taylor, "Statistical models of appearance for medical image analysis and computer vision," *SPIE*, vol. 4322, pp. 236–248, 2001.

[21] A. Hill, A. Thornham, and C. Taylor, "Model-based interpretation of 3-D medical images," *Proc. 4$^{\text{th}}$ British Machine Vision Conf.*, Sept. 1993, pp. 339–348, 1993.

[22] J. Yang and J. Duncan, "3D image segmentation of deformable objects with joint shape-intensity prior models using level sets," *Med. Image Anal.* vol. 8, pp. 285–294, 2004.

[23] D. Freedman, R. Radke, T. Zhang, Y. Jeong, D. Lovelock, and G. Chen, "Model-based segmentation of medical imagery by matching distributions," *IEEE Trans. on Medical Imaging*, vol. 24, no. 3, pp. 281–292, 2005.

[24] A. Foulonneau, P. Charbonnier, and F. Heitz, "Affine-invariant geometric shape priors for region-based active contours," *IEEE Trans. PAMI*, vol. 28, pp. 1352–1357, 2006.

[25] D. Adalsteinsson and J. Sethian, "A fast level set method for propagating interfaces," *J. Computational Physics*, vol. 118, pp. 269–277, 1995.

[26] M. Hassouna and A. Farag, "Robust centerline extraction framework using level sets," *In Proc. CVPR*, San Diego, USA, 2005, pp. 458-465.

[27] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comp. Vision*, vol. 60, pp. 91–110, 2004.

[28] G. Gimel'farb and D. Zhou, "Accurate identification of a Markov-Gibbs model for texture synthesis by bunch sampling," *Proc. CAIP*, Vienna, Austria, 2007, pp. 979-986.

[29] G. Gimel'farb, *Image textures and Gibbs random fields*, Kluwer Acad., Dordrecht, 1999.

[30] A. El-Baz, "Novel stochastic models for medical image analysis," *Ph.D. Thesis*, University of Louisville, Louisville, KY, August, 2006.