

CRF-driven Implicit Deformable Model

Gabriel Tsechpenakis
University of Miami

Dept. of Electrical and Computer Engineering
gavriil@miami.edu

Dimitris N. Metaxas
Rutgers University

Dept. of Computer Science
dnm@cs.rutgers.edu

Abstract

We present a topology independent solution for segmenting objects with texture patterns of any scale, using an implicit deformable model driven by Conditional Random Fields (CRFs). Our model integrates region and edge information as image driven terms, whereas the probabilistic shape and internal (smoothness) terms use representations similar to the level-set based methods. The evolution of the model is solved as a MAP estimation problem, where the target conditional probability is decomposed into the internal term and the image-driven term. For the later, we use discriminative CRFs in two scales, pixel- and patch-based, to obtain smooth probability fields based on the corresponding image features. The advantages and novelties of our approach are (i) the integration of CRFs with implicit deformable models in a tightly coupled scheme, (ii) the use of CRFs which avoids ambiguities in the probability fields, (iii) the handling of local feature variations by updating the model interior statistics and processing at different spatial scales, and (v) the independence from the topology. We demonstrate the performance of our method in a wide variety of images, from the zebra and cheetah examples to the left and right ventricles in cardiac images.

1. Introduction

Object segmentation plays a fundamental role in both computer vision and medical image analysis. A challenging task is to segment objects with textures of different scales, in the presence of clutter, complex backgrounds, and insufficiencies in the object boundaries. To address these problems, many segmentation approaches have been proposed. In this work we focus on a specific category of segmentation methods, namely the deformable models.

Parametric deformable models [9, 19, 2], or active contours, use parametric curves to represent the model shape. They start from an initial estimate for the boundary of the region of interest (ROI) and use image-driven forces to move the curve towards the desired boundaries, while their

internal (smoothness) forces preserve the curve's smoothness along its arc-length. The *traditional* parametric methods, use edges as image features to attract the curve towards the desired position, which makes them sensitive to noise, background complexity, and boundary edges' insufficiencies. This is the reason why other parametric methods have been proposed that use region-based features for the image-driven forces [18], or even combinations of both edge- and region-based forces [21]. The limitation of these methods is that they do not update the region statistics during the model evolution, so that local feature variations are difficult to be captured. A method that can be used for updating the region statistics was proposed in [4], where Particle Filters are used for vascular segmentation with active contours in medical images. In this method, boundary insufficiencies are successfully handled; on the other hand, this method is not integrated into a topology-independent deformable model and it not used yet for more complex textures.

Another category of deformable models are the implicit geometric models [15, 17, 1], which use the level-set based shape representation. In the Mumford-Shah formulation, the objective function to be minimized consists of implicit terms for the image-driven forces, and the boundary smoothness and length constraints. According to this formulation the optimal curve is the one that best approximates the image data, it is smooth locally and has the minimum length. In [17] a variational framework is proposed, integrating boundary and region-based information in PDEs that are implemented using a level-set approach. However, these frameworks assume piecewise or Gaussian intensity distributions within each partitioned image region, which limits their ability to capture intensity inhomogeneities and complex intensity distributions. Moreover, the high computational cost of the level-set based methods is widely known. To tackle these problems, another class of deformable models were proposed in [8], namely the Metamorphs. This framework uses both edge and texture information in a semi-parametric model, where the shape is represented similarly as in level-set methods. The interior texture is captured using a nonparametric kernel-based approx-

imation of the intensity probability density function (*pdf*), which is updated in every iteration. The deformations that the model can undergo are defined using Free Form Deformations (FFD).

Among the parametric and implicit deformable models, and apart from the traditional energy minimization, there are methods that convert the energy-driven model evolution into a maximum a posteriori (MAP) problem, using a probabilistic formulation. In the survey of [14], methods that use probabilistic formulations are described. The main drawback of most of these methods is that they use only edge information, although they are more robust than similar deterministic approaches. In the work of [6] the integration of probabilistic active contours with Markov Random Fields (MRFs) in a graphical framework was proposed to overcome the limitations of edge-based probabilistic active contours. Deformable models with MRFs were also used in [7] to achieve better (smoother) image likelihoods for the model evolution. In this work, although the MRFs were loosely integrated with the deformable model, the results show that the use of MRFs outperforms methods that do not use smoothing in the label field.

As an alternative to the MRFs, and to obtain better smoothing in the probability (or label) fields, Conditional Random Fields (CRFs) were introduced in computer vision [12]. Although CRFs were first used to label sequential data, extensions of them were used for image segmentation [10, 11, 5, 20]. The main advantage of CRFs compared to MRF-based segmentation (such as [13]), is that they handle the known label bias problem, described in [12], taking into account the pixel or region features in the smoothing of the probability field. In the work of [10] the discriminative version of CRFs was presented, to incorporate spatial neighborhood dependencies both in the labels and the image features, allowing the MAP inference for binary pixel classification. In the work of [5], CRFs were used in different spatial scales to capture the dependencies between image regions of multiple sizes. CRFs were also used in image sequences as in [20], where the dynamic CRF model was proposed to capture both temporal and spatial contextual information.

In this work we present a probabilistic implicit deformable model that combines the advantages of the above described segmentation approaches. In our method, (i) we use the shape representation of known level-set based approaches, to achieve topology independence, (ii) we integrate edge and region information, which is being updated during the model evolution, to handle local feature variations, (iii) we avoid the problem of getting trapped in local minima, which most of the energy minimization driven models suffer from, (iv) we exploit the superiority of CRFs compared to MRFs for image segmentation, coupling a CRF-based scheme with the deformable model, and (v) we

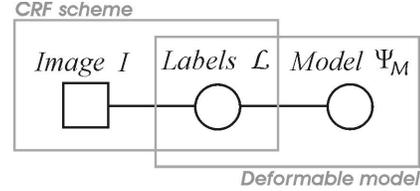


Figure 1. Graphical model for the integration of the CRF scheme in the deformable model

capture higher scale dependencies, using pixel- and patch-based CRFs. We use the two-scale CRF model in a tightly coupled framework with the deformable model, such that the external (image-driven) term of the deformable model eventually corresponds to the smooth probability field estimated by the CRF. We use a modified version of the discriminative CRFs presented in [10], where the MAP inference is computationally tractable using graph min-cut algorithms. Unlike methods such as the Active Shape Model (ASM) [3] and the Active Elastic Model [16], in this work we do not use shape priors.

The rest of this paper is as organized follows. In the next section we describe the formulation of our deformable model as a MAP problem. In 3 we describe the multi-scale CRF model that uses the discriminative CRF formulation. In 3.1 and 3.2 we describe the CRF formulation for two different spatial scales, namely the pixel- and patch-based CRFs, whereas in 3.3 we describe the integration of the two CRFs. In 4 we present our results and in 5 we give our conclusions and describe our future work.

2. Deformable Model Formulation

We use an implicit representation of the evolving curve and we follow a probabilistic formulation of the energy terms, namely the image-driven and the smoothness term. Using the simple graphical model of Fig. 1, we integrate the deformable model with a CRF scheme; then, the energy minimization is solved as a MAP problem.

The model \mathcal{M} defines two regions in the image domain Ω , namely the region $\mathcal{R}_{\mathcal{M}}$ enclosed by the model \mathcal{M} and the background $\Omega \setminus \mathcal{R}_{\mathcal{M}}$. The model is represented implicitly by its distance transform, *i.e.*, the signed distance function,

$$\Psi_{\mathcal{M}}(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in \mathcal{M} \\ +d(\mathbf{x}, \mathcal{M}), & \mathbf{x} \in \mathcal{R}_{\mathcal{M}} \\ -d(\mathbf{x}, \mathcal{M}), & \mathbf{x} \in \Omega \setminus \mathcal{R}_{\mathcal{M}} \end{cases}, \quad (1)$$

where $\mathbf{x} = (x, y)$ is the image pixel location in cartesian coordinates, and $d(\mathbf{x}, \mathcal{M}) = \min_{\mathbf{x}_{\mathcal{M}}} \|\mathbf{x} - \mathbf{x}_{\mathcal{M}}\|$ is the minimum Euclidean distance from the pixel location \mathbf{x} to the model consisting of the points $\mathbf{x}_{\mathcal{M}}$.

Our objective is formulated as a joint MAP estimation problem,

$$\langle \Psi_{\mathcal{M}}^*, \mathcal{L}^* \rangle = \arg \max_{(\Psi_{\mathcal{M}}, \mathcal{L})} P(\Psi_{\mathcal{M}}, \mathcal{L} | I), \quad (2)$$

where \mathcal{L} is the pixel labels and I is the given image. According to the model of Fig. 1, the probability $P(\Psi_{\mathcal{M}}, \mathcal{L}|I)$ is decomposed into,

$$\begin{aligned} P(\Psi_{\mathcal{M}}, \mathcal{L}|I) &\propto P(I|\mathcal{L}) \cdot P(\mathcal{L}|\Psi_{\mathcal{M}}) \cdot P(\Psi_{\mathcal{M}}) \\ &= P(\Psi_{\mathcal{M}}) \cdot P(I) \cdot P(\mathcal{L}|\Psi_{\mathcal{M}}) \cdot P(\mathcal{L}|I) \end{aligned} \quad (3)$$

The term $P(\Psi_{\mathcal{M}})$ corresponds to the model internal energy and the probability $P(I)$ is the image prior, assuming, without loss of generality, a gaussian distribution for the pixel intensities. The term $P(\mathcal{L}|\Psi_{\mathcal{M}})$ will be defined as a softmax function of $\Psi_{\mathcal{M}}$. Finally, the main contribution of this paper is the estimation of the probability $P(\mathcal{L}|I)$, which is estimated using our CRF-based approach.

In the following, we will describe the terms $P(\Psi_{\mathcal{M}})$, $P(I)$ and $P(\mathcal{L}|\Psi_{\mathcal{M}})$, and in section 3 we describe how we formulate the term $P(\mathcal{L}|I)$.

The image prior $P(I)$: For simplicity, and without loss of generality, we define the image prior $P(I)$ in terms of a gaussian distribution,

$$P(I(\mathbf{x}_i)) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-I(\mathbf{x}_i)^2/\sigma_0^2\right\} \quad (4)$$

One can use a nonparametric formulation for this prior.

The term $P(\mathcal{L}|\Psi_{\mathcal{M}})$: The term $P(\mathcal{L}|\Psi_{\mathcal{M}})$ is defined in a similar way as in [6]; for each image pixel \mathbf{x}_i , using the softmax function,

$$P(l_i|\Psi_{\mathcal{M}}) = \frac{1}{1 + \exp\{-\Psi_{\mathcal{M}}(\mathbf{x}_i)\}} \quad (5)$$

The model prior $P(\Psi_{\mathcal{M}})$: The term $P(\Psi_{\mathcal{M}})$ corresponds to the model internal energy, *i.e.*, the model smoothness. We define the model internal energy in terms of the area of the model interior, and the first derivative of the model distance transform,

$$\begin{aligned} E_{int}(\Psi_{\mathcal{M}}) = \\ \varepsilon_1 \mathcal{A}(\mathcal{R}_{\mathcal{M}}) + \varepsilon_2 \int \int_{\partial\mathcal{R}_{\mathcal{M}}} \|\nabla\Psi_{\mathcal{M}}(\mathbf{x})\| d\mathbf{x} \end{aligned} \quad (6)$$

where $\partial\mathcal{R}_{\mathcal{M}}$ denotes a narrow band around the model. The parameters $\varepsilon_1, \varepsilon_2$ are weighting constants, and $\mathcal{A}(\mathcal{R}_{\mathcal{M}})$ is the area of the model interior $\mathcal{R}_{\mathcal{M}}$, which is calculated as,

$$\mathcal{A}(\mathcal{R}_{\mathcal{M}}) = \int \int_{\Omega} \mathcal{H}(\Psi_{\mathcal{M}}(\mathbf{x})) d\mathbf{x}, \quad (7)$$

with \mathcal{H} being the step function: $\mathcal{H}(x) = 1, \forall x \geq 0$, and $\mathcal{H}(x) = 0, \forall x < 0$.

In a typical deformable model that follows an energy minimization approach, the first term of eq. (6) would force the model to a position where the area of its interior is the minimum, whereas the second term would enforce first-order smoothness. Similarly to the works of [14, 6], the

model internal energy can be written in a probabilistic manner using a gibbs prior as,

$$P_{int}(\Psi_{\mathcal{M}}) = (1/Z_{int}) \exp\{-E_{int}(\Psi_{\mathcal{M}})\}, \quad (8)$$

where Z_{int} is a constant.

From eq. (3), the remaining term $P(\mathcal{L}|I)$ is formulated using our CRF-based approach described in the following section.

3. CRF-based Image Term

We use a two-scale discriminative CRF formulation, where the scale refers to pixels and image regions (patches). In the following, we describe the general CRF framework that is applied to both scales.

Let $\mathcal{L} = \{l_i\}$ be the labels associated to the image sites $\mathbf{s} = \{s_i\}$. For the sake of generality, we use the term *sites* to refer to either the image pixels or the image patches. In our case, the sites (either the image pixels or the image patches) can have two labels, *i.e.*, belong to either the model interior ($\mathcal{R}_{\mathcal{M}}$) or the background ($\Omega \setminus \mathcal{R}_{\mathcal{M}}$).

The discriminative CRFs can estimate directly the labels distribution of the sites \mathbf{s} given an appropriate set of features of the image I as,

$p(\mathcal{L}|I) =$

$$\frac{1}{Z} \exp\left\{\sum_{i \in \mathcal{S}} \phi_a(l_i, I) + \sum_{i \in \mathcal{S}} \sum_{j \in N_i} \phi_i(l_i, l_j, I)\right\}, \quad (9)$$

where Z is a normalization constant, \mathcal{S} is the size of the set \mathbf{s} (the number of sites), and N_i is the set of neighbors of the site (l_i, s_i) .

The association potential $\phi_a(l_i, I)$ between a site label l_i and the observation set I is defined as the log probability,

$$\phi_a(l_i, I) = \log P(l_i|I) \quad (10)$$

The interaction potential $\phi_i(l_i, l_j, I)$ between neighboring site labels l_i and l_j given the observations I is,

$$\phi_i(l_i, l_j, I) = \frac{1}{z_i} \exp\left\{\frac{\delta(l_i - l_j)}{\sigma^2}\right\} \cdot \mathcal{F}(I), \quad (11)$$

where $\delta(x) = 1$, if $x = 0$ and $\delta(x) = 0$, if $x \neq 0$, \mathcal{F} is a function of the image features used to drive the deformable model, z_i is a normalization constant, and σ^2 controls the similarity between neighboring labels. Note that in common MRFs, similar definition is used for the interaction between neighboring sites with the difference that instead of the function $\mathcal{F}(I)$, a global constant (usually equal to one) is used.

The association potential *associates* the site's label with the observation in a discriminative manner. The interaction potential acts as a data-driven term that handles discontinuities, *i.e.*, forces smoothness in the probability field

when neighboring sites have different observations; otherwise, when the observations of neighboring sites *agree*, the interaction potential has no effect.

In our CRF-based approach, we first estimate the pixel-based and then the patch-based probability field. In both cases we use the above CRF formulation.

3.1. Pixel-based Probability Field

We will estimate the local probability field according to eq. (9), where *local* refers to the pixel-based CRF, the sites refer to the image pixels $\mathbf{X} = \{\mathbf{x}_i\}$, where \mathbf{x}_i is the cartesian coordinates on the image plane.

3.1.1 Pixel-based Association Potential of eq. (10)

We define the association potential of eq. (10) using two pixel-wise features, namely the pixel intensity and the texture of the region centered at a pixel. For the estimation of the probabilities below, we use training samples for the intensity and the texture of the ROI, and we update the region statistics using the model interior (from the initialization and the model interior at each iteration).

(a) Pixel intensity: Let $p_I(\mathbf{x}) = p(\mathbf{x} \in \mathcal{R}_{\mathcal{M}} | I(\mathcal{R}_{\mathcal{M}}))$ be the probability density function (*pdf*) of the intensity values at the locations \mathbf{x} corresponding to the model interior $\mathcal{R}_{\mathcal{M}}$. We approximate this *pdf* with a mixture of k gaussians,

$$p_I(\mathbf{x}) = \sum_{i=1}^k \omega_i \cdot g(\mu_i, \sigma_i^2) \quad (12)$$

where the parameters $\vartheta = [(\omega_i, \mu_i, \sigma_i^2), i = 1, \dots, k]$ are estimated using the EM algorithm. Then, the probability of a pixel label given its intensity value is,

$$P(l_i | I(\mathcal{R}_{\mathcal{M}})) = P_I(\mathbf{x}_i), \quad (13)$$

where $I(\mathcal{R}_{\mathcal{M}})$ denotes the intensity values in the model interior. The *pdf* of the intensity values in the model interior is learnt off-line using training samples for the desired ROI, and the parameters ϑ are updated during the model evolution.

(b) Multi-scale texture: For very pixel \mathbf{x}_i we estimate the probability of its neighborhood $N(\mathbf{x}_i)$ being consistent with the desired ROI, learned from both the training samples and the model interior.

We represent this probability with $P_T(\mathbf{x}_i) = P(N(\mathbf{x}_i) \in \mathcal{R}_{\mathcal{M}} | T(\mathcal{R}_{\mathcal{M}}))$, where $T(\mathcal{R}_{\mathcal{M}})$ denotes the ROI's texture. To estimate the texture of a neighborhood, we use the gabor filter responses for different scales (frequencies) and orientations. Under the assumption that the gabor responses are conditionally independent, the probability of a pixel label given its neighborhood's texture is given by,

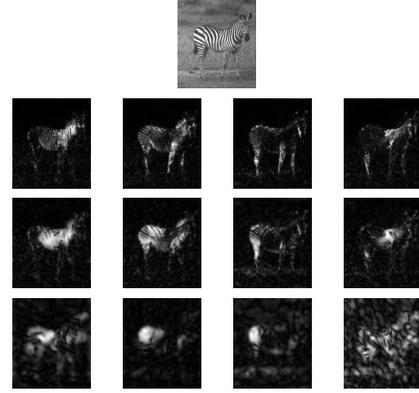


Figure 2. Gabor responses for three scales (rows) and four orientations (columns)

$$\begin{aligned} P(l_i | T(\mathcal{R}_{\mathcal{M}})) &= P_T(\mathbf{x}_i) \\ &= P(\bar{\xi}_i | \bar{\xi}_{\mathcal{M}}) = \prod_{s=1}^S \prod_{o=1}^O P(\bar{\xi}_i | \xi_{\mathcal{M}}^{(s,o)}) \\ &\propto \prod_{s=1}^S \prod_{o=1}^O P(\xi_i^{(s,o)} | \xi_{\mathcal{M}}^{(s,o)}), \end{aligned} \quad (14)$$

where S and O denote the the gabor filters scales and orientations, respectively, and $\bar{\xi}_{\mathcal{M}} = \{\xi_{\mathcal{M}}^{(s,o)}\}$ and $\bar{\xi}_i = \{\xi_i^{(s,o)}\}$ are the filter responses for the desired ROI and the i -th pixel neighborhood. Fig. 2 illustrates the gabor responses in three scales (rows) and four orientations (columns) for the zebra example.

(c) Integration of intensity and texture: The pixel-based association potential for a given image pixel, as defined in eq. (10), is the log conditional probability of the pixel label given the observation. Thus, according to the individual probabilities described above,

$$P_{pixel}(l_i | I) = P(l_i | I(\mathcal{R}_{\mathcal{M}})) \cdot P(l_i | T(\mathcal{R}_{\mathcal{M}})), \quad (15)$$

assuming that intensity and texture are independent. Note that estimating these probabilities independently has the same result as estimating association potentials for each feature separately, since

$$\begin{aligned} \phi_a^{local}(l_i, I) &= \log[P_I(\mathbf{x}_i) \cdot P_T(\mathbf{x}_i)] \\ &= \log P_I(\mathbf{x}_i) + \log P_T(\mathbf{x}_i) \\ &= \phi_a^{local}(l_i, I(\mathcal{R}_{\mathcal{M}})) + \phi_a^{local}(l_i, T(\mathcal{R}_{\mathcal{M}})) \end{aligned} \quad (16)$$

3.1.2 Pixel-based Interaction Potential of eq. (11)

We define the interaction potential of eq. (11) using the intensity and texture information. The goal is to define such a function $\mathcal{F}(I)$ that can describe the similarity between (neighboring) sites, based on the corresponding image features.

Let $f_I(\mathbf{x}_i, \mathbf{x}_j)$ and $f_T(\mathbf{x}_i, \mathbf{x}_j)$ be the similarity measures between two pixels \mathbf{x}_i and \mathbf{x}_j , based on the grayscale intensity, and the texture of the regions centered at those pixels.

The function $\mathcal{F}(I)$ is defined locally as,

$$\mathcal{F}^{local}(I) = \alpha f_I(\mathbf{x}_i, \mathbf{x}_j) + \beta f_T(\mathbf{x}_i, \mathbf{x}_j) \quad (17)$$

where α and β are constants. In our results, we used equal weights $\alpha = \beta = 1$.

(a) Pixel intensity: The simplest similarity measure between two pixels is the absolute difference between their intensities,

$$f_I(\mathbf{x}_i, \mathbf{x}_j) = \|I(\mathbf{x}_i) - I(\mathbf{x}_j)\| \quad (18)$$

This term is not sufficient for our purposes, since texture patterns of higher scales (textons with size bigger than a single pixel) include pixels of much different intensities.

(b) Multi-scale texture: Based on the texture information we described in 3.1.1, the similarity between two pixels is measured as the Bhattachayya distance between the *pdfs* of the gabor responses $\xi^{(s,o)}$ in the image regions centered at those pixels,

$$f_T(\mathbf{x}_i, \mathbf{x}_j) =$$

$$\sqrt{\sum_{s=1}^S \sum_{o=1}^O D_{Bhatt}^2[p(\xi_i^{(s,o)}) - p(\xi_j^{(s,o)})]}, \quad (19)$$

where the Bhattachayya distance between p_1 and p_2 is,

$$D_{Bhatt}(p_1, p_2) = -\log \left\{ \int [p_1(l)]^{\frac{1}{2}} [p_2(l)]^{\frac{1}{2}} dl \right\} \quad (20)$$

3.1.3 Pixel-based Probability field of eq. (9)

From the general CRF model of eq. (9), and using the association and interaction potentials described above, the probability field $p_{local}(\mathcal{L}|I)$ produced by the pixel-based CRF is estimated by replacing in eq. (9),

$$\begin{cases} \phi_a(l_i, I) = \phi_a^{local}(l_i, I) = \log P_{pixel}(l_i|I) \\ \phi_i(l_i, l_j, I) = \frac{1}{z_i} \exp \left\{ \frac{\delta(l_i - l_j)}{\sigma^2} \right\} \mathcal{F}^{local}(I) \\ S = \|\Omega\| \end{cases}, \quad (21)$$

where $\|\Omega\|$ is the size of the image, *i.e.*, the total number of pixels.

3.2. Patch-based Probability Field

From the probability field estimated by the pixel-based CRF, we go a step further and smooth the probability field to overcome discontinuities at a higher (spatial) scale.

Instead of using the *associations* between pixel labels-observations and *interactions* between pixels, we use the association and interaction potentials for image patches. Furthermore, we apply the smoothing process only for those

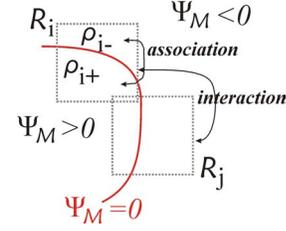


Figure 3. Patches around the model used as *sites* in the CRF.

image patches close to (around) the model to reduce the complexity. Instead of using features derived from the image, we use directly the pixel-based probability field. In this way, we now take into account the probability patterns within those patches. In Fig. 3, we illustrate the idea of smoothing the probability field using patches, which is explained below.

3.2.1 Patch-based Association Potential of eq. (10)

Let $\mathbf{R} = \{\mathcal{R}_r\}$ represent the set of the image patches \mathcal{R}_r centered at the image sites located in a narrow band around the model; obviously these patches are overlapping and cover a wider band around the model, depending on the patch size. Also, let $\mathbf{x}_{r,i}$ denote the image pixels within the r -th patch, $\mathcal{A}(\mathcal{R}_r)$ be the area of the patch, and i index the pixel locations within the patch. To determine whether \mathcal{R}_r is entirely in the interior or exterior of the model, we define the parameter,

$$A_r = \text{sign} \left\{ \mathcal{A}(\mathcal{R}_r) - \left\| \int \int_{\mathcal{R}_r} \text{sign}(\Psi_{\mathcal{M}}(\mathbf{x})) d\mathbf{x} \right\| \right\}, \quad (22)$$

where $\text{sign}(x) = 1$, if $x > 0$, $\text{sign}(x) = -1$, if $x < 0$ and $\text{sign}(x) = 0$, if $x = 0$. The above expression is equal to zero if $\Psi_{\mathcal{M}}(\mathbf{x}_{r,i}) > 0, \forall \mathbf{x}_{r,i} \in \mathcal{R}_r$, or $\Psi_{\mathcal{M}}(\mathbf{x}_{r,i}) < 0, \forall \mathbf{x}_{r,i} \in \mathcal{R}_r$. Also, if a patch shares image pixels from both the model interior and exterior, it is $A_r = 1$.

As mentioned above, we are interested in those patches where $A_r = 1$, *i.e.*, for the patches where there is at least one pair $(\mathbf{x}_{r,i}, \mathbf{x}_{r,j}), i \neq j$, that satisfies the condition $\text{sign}\{\Psi_{\mathcal{M}}(\mathbf{x}_{r,i})\} \neq \text{sign}\{\Psi_{\mathcal{M}}(\mathbf{x}_{r,j})\}$.

The probability of a patch with $A_r = 1$ belonging to the model interior is defined in terms of a gibbs prior, which associates the probability values of the pixels in the patch where $\Psi_{\mathcal{M}} < 0$, with the probability values of the pixels where $\Psi_{\mathcal{M}} \geq 0$.

More specifically, let P_{local} be the probability field estimated from the pixel-based CRF, and ρ_{r+}, ρ_{r-} be the distributions of the probability values inside the r -th patch, where $\Psi_{\mathcal{M}} \geq 0$ and $\Psi_{\mathcal{M}} < 0$ respectively (Fig 3). Then the Bhattachayya distance between ρ_{r+} and ρ_{r-} indicates how similar the probability patterns of these two intra-patch segments are. Thus, the probability of the entire patch \mathcal{R}_r

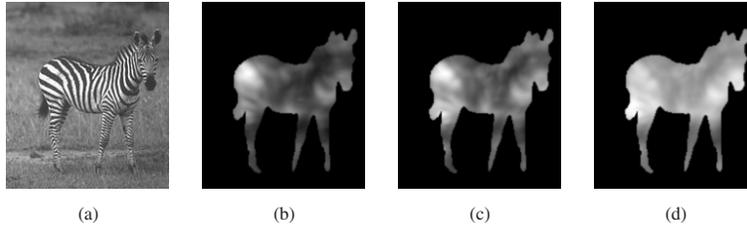


Figure 4. Comparison between MRF, the pixel-based CRF and the two-scale CRF.



Figure 5. Model dynamics for the zebra example.

final result are shown in a variety of images. All results are shown with red lines on the object boundaries.

In Fig. 5 we illustrate six indicative moments of the model evolution in the zebra example. The ROI texture samples we used in our pixel-based CRF scheme include the zebra-stripe pattern as well as samples from tails and noses. In Fig. 6(a) we illustrate a similar case, namely the cheetah example. The segmentation of the bird in Fig. 6(b) is another case where the object of interest has different texture patterns and the background is highly cluttered; in this case we used different training samples for all the texture patterns of the bird. In the submitted video ‘*bird-init-only.mpg*’ we show the segmentation result using only the region statistics of the model interior and no training samples; in this case we undersegment the bird.

For the lizard examples of Fig. 6(c) and (d), we show the performance of our method in cases where there is a very small difference between the ROI and the background textures and intensities. For the example of Fig. 6(c), the submitted video ‘*lizard-init-only.mpg*’ shows the results of the lizard segmentation using only the model interior statistics (without using training samples for the intensity and texture); in this case we undersegment the lizard.

In Fig. 6(e), although the two fish are successfully segmented from the background, the boundary between them is not found, since we do not use shape priors in our method (see future work).

In Fig. 6(f) and (g) we illustrate the performance of our method for the segmentation of the left and right ventricle in cardiac MRI images; the second image is a tagged MRI where the texture pattern is of higher scale. Although we used different intensity and texture training samples for these two cases, the two segmentation results are exactly the same. Finally, in Figs. 6(h) and (i), we illustrate our results

for local texture variations. For the example of Fig. 6(h) we used training samples that include the dark regions between the peppers; in the submitted video ‘*peppers-only.mpg*’ we show the same example where the training samples did not include those dark regions. For the example of Fig. 6(i) we only used the model interior statistics and no training samples.

5. Conclusions and Future Work

We presented a robust solution for segmenting objects with texture patterns of any scale, using an implicit deformable model driven by Conditional Random Fields (CRFs), in a tightly coupled framework. The deformable model is integrated with the CRF scheme using a simple graphical model, and the evolution is solved as a MAP estimation problem. Our CRF scheme is based on the discriminative version of the CRFs and it is used in two scales, namely pixel- and patch-based. In this way we obtain smooth probability fields based different image features, namely the pixel intensity and the multi-scale texture. We avoid ambiguities in the estimation of the probability field with our two-scale CRF scheme, and we handle local feature variations by updating the statistics of our model interior. Our results show that our method performs successfully in a wide variety of images.

We leave for future work the introduction of shape priors into our framework.

References

- [1] V. Caselles, R. Kimmel, and G. Sapiro, ‘Geodesic active contours,’ *IEEE Int’l Conference on Computer Vision (ICCV95)*, 1995. 1
- [2] L. D. Cohen and I. Cohen, ‘Finite-element Methods for Active Contour Models and Balloons for 2-D and 3-D Images,’

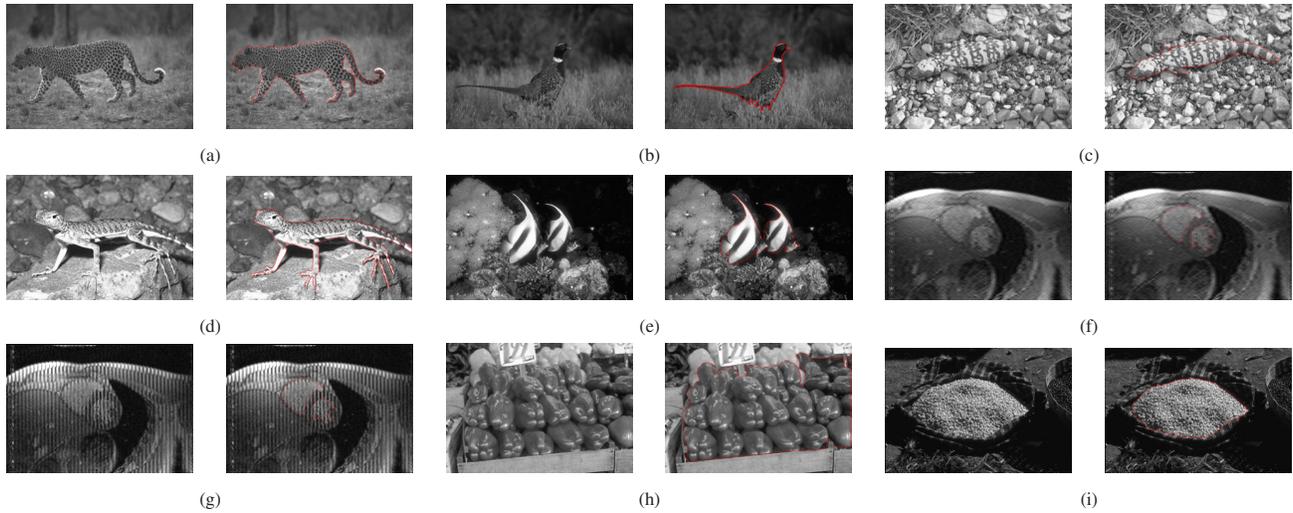


Figure 6. Segmentation examples for different objects.

- IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:1131-1147, 1993. 1
- [3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, 'Active shape models - their training and application,' *Computer Vision and Image Understanding*, 61(1):38-59, 1995. 2
- [4] C. Florin, J. Williams, and N. Paragios, 'Globally Optimal Active Contours, Sequential Monte Carlo and On-line Learning for Vessel Segmentation,' *European Conference on Computer Vision (ECCV06)*, 2006. 1
- [5] X. He, R. Zemel, and M. Carreira-Perpinan, 'Multiscale Conditional Random Fields for Image Labeling,' *IEEE Conference on Computer Vision and Pattern Recognition (CVPR04)*, 2004. 2
- [6] R. Huang, V. Pavlovic, and D. Metaxas, 'A Graphical Model Framework for Coupling MRFs and Deformable Models,' *IEEE Conference on Computer Vision and Pattern Recognition (CVPR04)*, 2004. 2, 3
- [7] X. Huang, Z. Qian, R. Huang, and D. Metaxas, 'Deformable Model Based Textured Object Segmentation,' *4th Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR05)*, 2005. 2
- [8] X. Huang, D. Metaxas, and T. Chen, 'Metamorphs: Deformable Shape and Texture Models,' *IEEE Conference on Computer Vision and Pattern Recognition (CVPR04)*, 2004. 1
- [9] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int'l Journal of Computer Vision*, 1:321-331, 1987. 1
- [10] S. Kumar and M. Hebert, 'Discriminative Fields for Modeling Spatial Dependencies in Natural Images,' *Advances in Neural Information Processing Systems, NIPS 16*, 2004. 2
- [11] S. Kumar and M. Hebert, 'Discriminative Random Fields: A Discriminative Framework for Contextual Interaction in Classification,' *IEEE International Conference on Computer Vision (ICCV03)*, 2003. 2
- [12] J. Lafferty, A. McCallum, and F. Pereira, 'Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data,' *Eighteenth International Conference on Machine Learning (ICML01)*, 2001. 2
- [13] S. Z. Li, 'Modeling Image Analysis Problems Using Markov Random Fields,' *Stochastic Processes: Modeling and Simulation*, Vol. 20 Handbook of Statistics, C.R. Rao and D.N. Shanbhag (ed), Elsevier Science. 2003. 2
- [14] T. McInerney and D. Terzopoulos, 'Deformable Models in Medical Image Analysis: A Survey,' *Medical Image Analysis*, 1(2), 1996. 2, 3
- [15] D. Mumford and J. Shah, 'Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems,' *Communications on Pure and Applied Mathematics*, 42(5):577-685, 1989. 1
- [16] X. Papademetris, E. T. Onat, A. J. Sinusas, D. P. Dione, R. T. Constable, and J. S. Duncan, 'The Active Elastic Model,' *Information Processing in Medical Imaging (IPMI01)*, 2001. 2
- [17] N. Paragios and R. Deriche, 'Geodesic Active Regions and Level Set Methods for Supervised Texture Segmentation,' *Int'l Journal of Computer Vision*, 46(3):223-247, 2002. 1
- [18] R. Ronfard, 'Region-based strategies for active contour models,' *International Journal of Computer Vision*, 13(2):229-251, 1994. 1
- [19] L. H. Staib and J. S. Duncan, "Boundary Finding with Parametrically Deformable Models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(11):1061-1075, 1992. 1
- [20] Y. Wang and Q. Ji, 'A Dynamic Conditional Random Field Model for Object Segmentation in Image Sequences,' *IEEE Conference on Computer Vision and Pattern Recognition (CVPR05)*, 2005. 2
- [21] S. Zhu and A. Yuille, 'Region Competition: Unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation,' *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884-900, 1996. 1