Level Set based Segmentation using Local Feature Distribution

Xianghua Xie

Department of Computer Science, Swansea University, UK.
x.xie@swansea.ac.uk

Abstract

We propose a level set based framework to segment textured images. The snake deforms in the image domain in searching for object boundaries by minimizing an energy functional, which is defined based on dynamically selected local distribution of orientation invariant features. We also explore the user initialization to simplify the segmentation and improve accuracy. Experimental results on both synthetic and real data show significant improvements compared to direct modeling of filtering responses or piecewise constant modeling.

1 Introduction

Region based active contours have been increasingly used in analyzing textured images, e.g. [8, 3]. Among many others, filtering is a popular approach in deriving texture features for contour based segmentation. For example, in [8] the Gabor filter responses are used to measure the difference between pixels in a piecewise constant model. However, it largely ignores the spatial distribution among local filtering coefficients and this direct comparison of filter responses is error prone since the responses can be misaligned due to the anisotropic nature of most of the filters. Handling large dimensional filtering responses can also be difficult. Moreover, it is challenging to deal with textural variations, for example, due to rotation or view point changes, since most of the filters are orientation sensitive. The separation of background and foreground has been popularly achieved through modeling of global distribution, e.g. mixture of Gaussian modeling [6, 2], or fitting of local distributions, e.g. piecewise constant assumption [1] as used in [8, 7, 5].

In this paper, we propose a level set based segmentation using a regional force based on nonparametric representations of local invariant features. These features are dynamically selected as histograms which is convenient to differentiate in minimizing the energy functional and more robust towards texture variation and inhomogeneity. Unlike [5] where only image intensity values are used, we advocate the use of filtering responses to deal with highly textured images. Prior knowledge extracted from user initialization can also be conveniently incorporated in the model to achieve efficient segmentation.

2 Proposed Approach

Briefly, we first extract compact and rotation invariant filtering responses from the image. Their local distributions at every pixel, known as local spectral histograms, are then collected. The optimal bin size for these histograms are automatically derived. An energy minimization problem is then formulated by fitting two spectral histograms, one of which is used to approximate the foreground region and the other for the background. To reduce the ambiguities among local features, a subset of filtering responses can be dynamically selected while performing segmentation. Finally, we present semisupervised segmentation based on initialization to demonstrate its flexibility and efficiency.

2.1 Filters and Feature Selection

Filter bank based approaches have been very popular since they can analyze textures in arbitrary orientations and scales. However, they often result in high dimensional feature space which can be difficult to handle for certain applications. Unlike image classification, in snake based segmentation, we may not have enough features extracted from a single image to populate the high dimensional feature space in order to accurately estimate the underlying distributions. Moreover, there are usually significant amount of redundant information among the filtering responses. It is also worth noting that object in the scene may have inhomogeneous textures due to, for example, perspective projection. This inhomogeneity will exhibit nonuniform responses after applying directional filters. Rotation invariance is thus desirable in such circumstance. Here, we use a set of directional filters derived from first and second order derivatives of an anisotropic Gaussian function, i.e. $\sigma_y = 3\sigma_x$. Each of these two types of filters are rotated to be uniformly spaced in six different directions.

This process is repeated in three progressive scales to produce a total of thirty six directional filters. In the interest of reducing feature space dimension and achieving rotation invariant, we follow [10] to condense the filter responses by collecting only the maximum filter response across all the six orientations, i.e. those thirty six directional filter responses are reduced to six.

2.2 Local Spectral Histogram

The filtering responses can be directly used to drive the active contours, e.g. [8]. However, we can further incorporate local spatial dependency of filtering responses by computing the marginal distributions over a local window. Thus, it captures local pixel dependency through filtering and global patterns through histograms. Misaligning of filter responses due to inhomogeneity of filter responses can be a serious problem for direct approaches. Using local spectral histogram further enhances our model in dealing with texture inhomogeneity and helps to produce more coherent segmentation. Let $\mathbf{W}$ denote a local window and $\mathbf{W}^{(\alpha)}(\mathbf{x})$ a maximum filter response patch centered at $\mathbf{x}$, where $\alpha = 1, 2, ..., 6$. Thus, for $\mathbf{W}^{(\alpha)}$ the histogram is defined as [4]:

$$P_{\mathbf{W}}^{(\alpha)}(z_1, z_2) = \sum_{\mathbf{x} \in \mathbf{W}} \int_{z_1}^{z_2} \delta(z - \mathbf{W}^{(\alpha)}(\mathbf{x}))dz, \quad (1)$$

where z_1 and z_2 specify the range of the bin. The spectral histogram is then defined as $P_{\mathbf{W}} = \frac{1}{\mathbf{W}} \left(P_{\mathbf{W}}^{(1)}, P_{\mathbf{W}}^{(2)}, ..., P_{\mathbf{W}}^{(6)} \right)$.

2.3 Deducing Optimal Bin Size

Automatically deducing suitable histogram bin size is desirable because not only it reduces the risk of large fluctuation or poor representation due to too small or too large bin size, but also avoids the practical issues associated with manual parameter tunning. We follow [9] to estimate the optimal bin size. Let Δ denote the bin size. The expected frequency for $s \in [0, \Delta]$ is $\theta = \frac{1}{\Delta} \int_0^{\Delta} \lambda_s ds$, where λ_s is the underlying true frequency which is unknown. The goodness of fit of the estimated $\hat{\lambda}_s$ to λ_s is measured according to mean integrated squared error: MISE $= \frac{1}{\Delta} \int_0^{\Delta} \langle E(\hat{\theta} - \lambda_s)^2 \rangle ds$, where E denotes expectation and $\hat{\theta}_i \equiv k_i/\Delta$ (k_i is the frequency count for ith bin). The associated cost function is then defined as: $\mathcal{O}(\Delta) = \text{MISE} - \frac{1}{\Delta} \int_0^{\Delta} \langle (\lambda_s - \langle \theta \rangle)^2 \rangle ds$. By assuming the number of events counted in each bin obeys a Poisson distribution, the cost function can be written as:

$$\mathcal{O}(\Delta) = \frac{2}{\Delta} \langle E\hat{\theta} \rangle - \langle E(\hat{\theta} - \langle E\hat{\theta} \rangle)^2 \rangle. \quad (2)$$

The optimal bin size thus is obtained by minimizing the above cost function, i.e. $\hat{\Delta} = \arg\min_{\Delta} \mathcal{O}(\Delta)$.

Thus, local spectral histograms at each pixel position are computed and combined together for those in the same channel to produce a global representation, i.e. mean local spectral histogram, followed by this optimal bin size selection.

2.4 Unsupervised Snake Segmentation

The snake segmentation can be viewed as a foreground-background partition problem. The snake evolves in the image domain, attempting to minimizing the feature similarity for those inside and outside the contours. Meanwhile, it tries to minimize the feature difference for those that belong to the same region. Thus, we can formulate our snake based on the piecewise constant model [1]. Let Ω be the image domain, Λ_+ denote the regions inside the snake (foreground) and Λ_- those outside the snake (background). The snake segmentation can be achieved by solving the following energy minimization problem:

$$\inf_{\Lambda_+} \mathcal{E}(\Lambda_+) = \beta\mathcal{L}(\Lambda_+) + \int_{\Lambda_+} \mathcal{D}(P(\mathbf{x}), P_+)d\mathbf{x}$$
$$+ \int_{\Lambda_-} \mathcal{D}(P(\mathbf{x}), P_-)d\mathbf{x}, \quad (3)$$

where β is a constant, $\mathcal{L}$ denote length, $\mathcal{D}$ is the metric which measures the difference between two histograms, and P_+ and P_- are the foreground and background mean local spectral histograms to be determined. The first term is the length minimization term which regularize the contour. The next two are data fitting terms.

The Wasserstein distance (aka. earth mover's distance) is used to compute the distance between two normalized spectral histograms since it is a true metric (unlike χ^2). Thus, the distance between two histograms (P_a and P_b) is defined as:

$$\mathcal{D}(P_a, P_b) = \int_T |F_a(y) - F_b(y)|dy, \quad (4)$$

where T denoted the range of the histogram bins, and F_a and F_b are cumulative distributions of P_a and P_b, respectively.

The level set method is implemented to solve this energy minimization problem so that topological changes, such as merging and splitting, can be effectively handled. Let ϕ denote the level set function. The foreground is identified as $\Lambda_+ = \{\mathbf{x} \in \Omega : \phi(\mathbf{x}) > 0\}$, which can be computed as $\int_\Omega \mathcal{H}(\phi)d\mathbf{x}$ where $\mathcal{H}$ is the Heaviside function. The level set formulation can be expressed as:

$$\inf_{\Lambda_+} \mathcal{E}(\Lambda_+) = \beta \int_\Omega |\nabla\mathcal{H}(\phi)|d\mathbf{x} + \int_\Omega \mathcal{D}(P(\mathbf{x}), P_+)\mathcal{H}(\phi)d\mathbf{x}$$
$$+ \int_\Omega \mathcal{D}(P(\mathbf{x}), P_-)(1 - \mathcal{H})(\phi)d\mathbf{x} \quad (5)$$
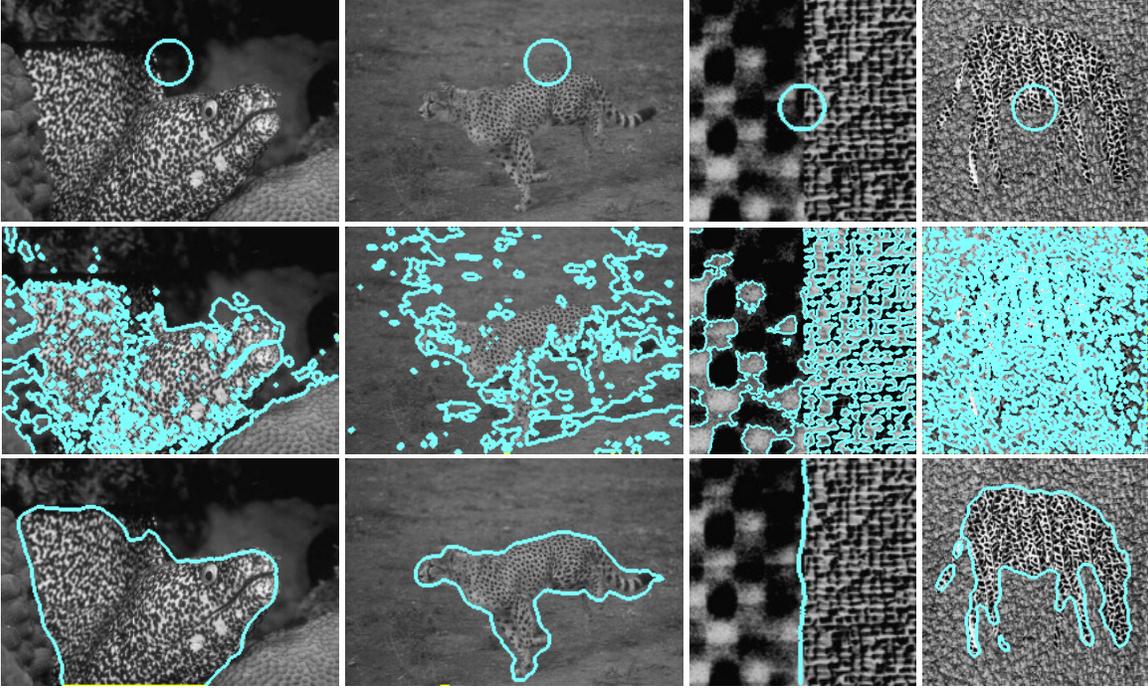
Figure 1. First row: initializations; second row: CV results; third row: proposed unsupervised.

Thus, minimizing $\mathcal{E}$ with respect to ϕ gives us the following partial differential equation:

$$\frac{\partial \phi}{\partial t} = \delta(\phi)\Big[\beta\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|}\right) - \int_T |F_{\mathbf{x}}(y) - F_+(y)|dy$$
$$+ \int_T |F_{\mathbf{x}}(y) - F_-(y)|dy\Big], \qquad (6)$$

where $\delta(x) = \frac{d}{dx}\mathcal{H}(x)$, F_+ and F_- are the spectral cumulative histogram inside and outside the contours, respectively. The minimization process thus moves the contours towards object boundaries through competing pixels by measuring the similarity of local cumulative spectral histogram with those inside and outside current foreground.

2.5 Dynamic Optimal Response Selection

Some of the maximum filter responses are inevitably more discriminative than the others. Thus, it is desirable to use a subset of the local spectral histograms, $P^{(\alpha)}, \alpha = 1, 2, ..., 6$, in solving (6). We empirically choose three out of the six spectral histograms. However, since P_+ and P_- are unknown, we estimate the optimal filter set by choosing those local spectral histograms that maximize the distance between P_+ and P_-. This estimation dynamically changes as the active contour evolves itself, and reduces the ambiguities between foreground and background, as well as further reduces feature space dimension.

2.6 Semisupervised Segmentation

When the prior knowledge of the region of interest is available, it is useful to include them in segmentation to achieve accurate results. The proposed method can be conveniently adapted to incorporate a reference local spectral histogram, obtained at the training stage, into the segmentation by forcing the foreground local spectral histogram distribution close to the reference histogram. However, extensive prior knowledge is not always available, particularly when segmenting single images. Here, we explore the user interaction of placing the initial contour. By placing the initial contour inside the region of interest, we can efficiently extract its optimal local spectral histogram as our reference histogram, i.e. $P_r = \frac{\int_\Omega \mathcal{H}(\phi_0)P'(\mathbf{x})d\mathbf{x}}{\int_\Omega \mathcal{H}(\phi_0)d\mathbf{x}}$, where ϕ_0 is the initial level set and P' is the optimal local spectral histogram. Let F_r denote the cumulative histogram of P_r. Thus, the semi-supervised segmentation can be written as (cf. (6)):

$$\frac{\partial \phi}{\partial t} = \delta(\phi)\Big[\beta\nabla \cdot \left(\frac{\nabla \phi}{|\nabla \phi|}\right) - \int_T |F_{\mathbf{x}}(y) - F_r(y)|dy$$
$$+ \int_T |F_{\mathbf{x}}(y) - F_-(y)|dy\Big]. \qquad (7)$$

The segmentation is simplified since it only needs to find the fitting function F_-. This approach is particularly useful when dealing with background texture with large variations.
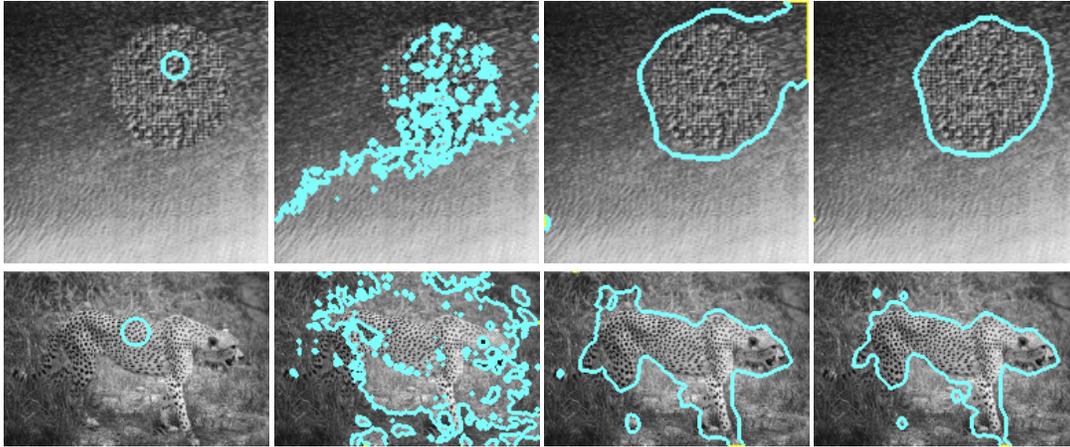
Figure 2. First column: initializations; second column: CV results; third column: proposed unsupervised method; fourth column: proposed semisupervised method.
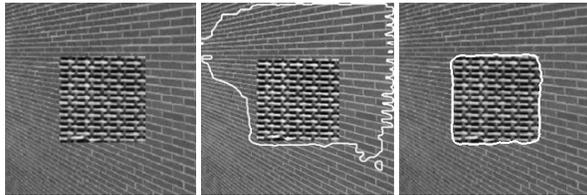
Figure 3. From left: a synthetic texture collage which contains an inhomogeneous background due to orientation and scale changes; segmentation result obtained using the CV model based on Gabor features [8] (result reported in [7]); and the proposed method.

3   Experimental Results

The proposed methods have been tested on both synthetic and real world images. Fig. 1 shows some typical example results obtained using unsupervised segmentation compared to the Chan-Vese (CV) method [1]. In all cases, the proposed method achieved significantly better segmentation results, only missing some very fine and thin structures.

Fig. 2 provides two examples of the proposed method using semisupervised approach. The initial snakes only cover a very small part of the object of interest. The images contain large variations in the background and are difficult to segment. The CV model failed to provide meaningful segmentation, whereas the proposed method managed to achieve reasonable results. It also demonstrates the advantage of incorporating even very limited prior knowledge.

Fig. 3 shows the proposed method performs significantly better than that directly using filter responses [8] when dealing with inhomogeneous textures.

4   Conclusions

In this paper, we introduced a region based snake segmentation framework which is based on nonparametric representations of condensed local filtering responses. Local spectral histograms are found to be effective and convenient in energy minimization formulation. The experimental studies showed some very promising results on various synthetic and real images.

References

[1] T. Chan and L. Vese. Active contours without edges. *IEEE T-IP*, 10(2):266–277, 2001.

[2] Y. He, Y. Luo, and D. Hu. Unsupervised texture segmentation via applying geodesic active regions to Gaborian feature space. *IEEE T-ECT*, pages 272–275, 2004.

[3] N. Houhou and J. Thiran. Fast texture segmentation model based on the shape operator and active contour. In *IEEE CVPR*, pages 1–8, 2008.

[4] X. Liu and D. Wang. Image and texture segmentation using local spectral histograms. *IEEE T-IP*, 15(10):3066–3077, 2006.

[5] K. Ni, X. Bresson, T. Chan, and S. Esedoglu. Local histogram based segmentation using the wasserstein distance. *IJCV*, 84:97–111, 2009.

[6] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *IJCV*, 46(3):223–247, 2002.

[7] C. Sagiv, N. Sochen, and I. Zeevi. Integrated active contours for texture segmentation. *IEEE T-IP*, 15(6):1633–1645, 2006.

[8] B. Sandberg, T. Chan, and L. Vese. A level-set and gabor-based active contour algorithm for segmenting textured images. Technical Report 39, UCLA, 2002.

[9] H. Shimazaki and S. Shinomoto. A method for selecting the bin size of a time histogram. *Neural Computation*, 19(6):1503–1527, 2007.

[10] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *ECCV*, pages 255–271, 2002.