GRAPH CUT SEGMENTATION WITH NONLINEAR SHAPE PRIORS

James Malcolm *Yogesh Rathi* *Allen Tannenbaum*

Georgia Tech
Atlanta, Georgia

ABSTRACT

Graph cut image segmentation with intensity information alone is prone to fail for objects with weak edges, in clutter, or under occlusion. Existing methods to incorporate shape are often too restrictive for highly varied shapes, use a single fixed shape which may be prone to misalignment, or are computationally intensive.

In this note we show how highly variable nonlinear shape priors learned from training sets can be added to existing iterative graph cut methods for accurate and efficient segmentation of such objects. Using kernel principle component analysis, we demonstrate how a shape projection pre-image can induce an iteratively refined shape prior in a Bayesian manner. Examples of natural imagery show that both single-pass and iterative segmentation fail without such shape information.

Index Terms— Image segmentation, graph cuts, shape priors, kernel PCA

1. INTRODUCTION

This note addresses the problem of semi-automatic image segmentation where the object of interest may have weak edges, lie next to objects of similar intensity distributions, or is partially occluded. Often in such problems, the segmentation task is formulated as one of energy minimization.

Recently, graph cut techniques have received considerable attention as a robust method for global energy minimization. Popularized for image segmentation by [1], graph cut techniques have found application throughout the vision community. See [2] for a history and survey of the field.

Despite its widespread application to image segmentation, in situations such as this, the iterative graph cut technique may not capture weak edges, may leak out of the object of interest, or be unable to capture occluded portions of the object. Similar to how intensity information is iteratively refined, incorporating prior shape information prevents such failures; however, incorporating such information into the graph cut method is an ongoing area of research which this paper extends.

In this work, we use kernel PCA [3] to learn a statistical model of relevant shapes. From a user-initialized segmentation, the algorithm proceeds iteratively. At each iteration the previous segmentation is used to obtain a projection in the learned shape space. In addition, the intensity histograms are recomputed given the previously segmented regions. These priors–the shape and histograms–are then used in a Bayesian formulation to perform segmentation via the graph cut technique.

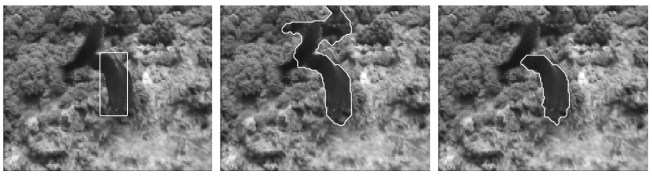

Fig. 1. Octopus near fish of similar intensity: initial, without shape, with shape *(left to right)*.

2. GRAPH CUTS

Taking advantage of efficient algorithms for global min-cut solutions, we cast the energy-based image segmentation problem in a graph structure of which the min-cut corresponds to a globally optimal segmentation.

Evaluated for an object/background pixel assignment $A = \{a_p : a_p \in \{\mathcal{O}, \mathcal{B}\}, p \in \mathcal{I}\}$, such energies are designed as a data dependent term and a smoothness term. The data dependent term evaluates the penalty for assigning a particular pixel to a given region. The smoothness term evaluates the penalty for assigning two neighboring pixels to different regions, i.e. a boundary discontinuity. These two terms are thought of as a region-based term and a boundary term, often weighted by $0 \leq \lambda$ for relative influence:

$$E(A) = \sum_{p \in \mathcal{I}} R_p(a_p) + \lambda \sum_{\substack{(p,q) \in \mathcal{N} \\ a_p \neq a_q}} B_{(p,q)} \tag{1}$$

where $\mathcal{I}$ represents all image pixels, $\mathcal{N}$ all unordered neighborhood pixel pairs.

To construct the graph representing this energy, each pixel is considered as a graph node in addition to two nodes representing object and background. The data dependent term is realized by connecting each pixel to both the object and background nodes with non-negative edge weights $R_p(\mathcal{O})$ and $R_p(\mathcal{B})$ representing the likelihood of object and background region presence at pixel p. Lastly, the smoothness term is realized by connecting each pairwise combination of neighboring pixels (p, q) with a non-negative edge weight determined by a penalty for boundary discontinuity, $B_{(p,q)}$. Notice that, since the min-cut sums only along the boundary, the boundary condition of $a_p \neq a_q$ in (1) may be ignored and every pair of neighboring pixels may be connected with edge weight $B_{(p,q)}$. The min-cut of the weighted graph represents the segmentation that best separates the object from its background. See [1] for more details.

Typical applications of graph cuts to image segmentation differ only in the definitions of R_p and $B_{(p,q)}$. For example, in [1] the negative log-likelihood of a pixel's fit into user-initialized intensity histograms is used in the regional term while intensity contrast is

used in the boundary term:

$$R_p(\mathcal{O}) = -\ln P(\mathcal{I}_p|\mathcal{O}) \qquad R_p(\mathcal{B}) = -\ln P(\mathcal{I}_p|\mathcal{B}) \quad (2)$$

$$B_{(p,q)} = \exp(\tfrac{-\|\mathcal{I}_p - \mathcal{I}_q\|^2}{2\sigma^2}) \tfrac{1}{\|p-q\|} \quad (3)$$

where σ is a user-defined parameter and $\|p-q\|$ is the Euclidean pixel distance for normalizing among edges of different length.

3. RELATED WORK

This note further integrates two areas of research, namely learned shape priors and graph cut segmentation techniques.

Learned shape priors have been used in segmentation techniques in a variety of ways. From a training set of segmented objects, such techniques build a statistical shape model using a particular representation and then attempt to find the segmentation that best fits the shape model [4]. This paper uses nonlinear kernel principle component analysis to form a shape model much like the approach taken in level set methods [5]. This approach easily captures the variance in the training set allowing for a highly deformable shape model.

Shape priors have been incorporated into graph cut techniques in a variety of ways. For example, the authors of [6] propose a novel method for segmenting a particular class of shapes that they define as compact. Here they modify the boundary term $B_{(p,q)}$ to be prohibitively large if the assignment violates the rules for maintaining a compact shape. However, the class of shapes defined as compact may be quite restrictive in applications involving highly variable shapes.

Another similar approach was that of [7] where segmentation was performed iteratively, at each iteration fitting an ellipse to the current segmentation defining a shape prior and band to segment within; however, like the compact shape approach, this does not generalize to highly variable shapes. Also, this method incorporates the shape prior as an additive fixed weight, not in the sense of a probabilistic Bayesian prior.

The iterative re-estimation of model parameters has seen application as a way of lessening the need for user interaction. In [8] segmentation proceeds iteratively, at each iteration the color histograms are updated from the regions defined by the last segmentation (or user initialization) and the segmentation performed once again with these new histograms. Similarly, [7] refits an ellipse to the last segmentation defining a new shape prior.

Seeking to capture more arbitrary shapes, the authors of [9] use a distance function in the smoothness term favoring boundary placement that is close to the shape outline. Determining the shape to segment and fixing its alignment before performing the graph cut, this method seems quite sensitive to those initial estimates. Further, since the shape prior comes into play in the boundary term, in a sense, this has a local effect and may again be sensitive to misalignment.

In a more generic approach to shape, [10] uses a Markov random field representation where the latent shape model variables are integrated out via expectation maximization. While shape information is utilized in a principled Bayesian manner, this approach is computationally intensive requiring a separate energy minimization and repeated sampling from the shape model.

A different method of incorporating shape involves the concept of flux across the segmented surface. Utilizing an aligned and fixed distance function, [11] seeks a segmentation which additionally minimizes the divergence between surface normals of the shape and segmentation. Aside from the complicated numerical scheme, the segmentation is sensitive to initial alignment; highly deformed objects significantly increase flux across the surface when slightly misaligned.

Other work such as [12] constrains shape, not with a statistical prior, but by narrowing the region of possible segmentations to a band around the user initialized contour. Such methods do not capture occluded regions.

4. KERNEL PRINCIPLE COMPONENT ANALYSIS

Kernel methods, and in particular, kernel principle component analysis (PCA) has been a focus of research in the pattern recognition community [3]. The basic idea behind these methods is to map the data from the input space $\mathbb{R}^n$ to a feature space $\mathcal{H}$ via some nonlinear map $\varphi : \mathbb{R}^n \rightarrow \mathcal{H}$, and then apply a linear method in $\mathcal{H}$ to do further analysis. Kernel PCA [13] is a nonlinear feature extractor where linear PCA is performed in the feature space $\mathcal{H}$ which is equivalent to doing nonlinear PCA in the input space $\mathbb{R}^n$. Because it is able to capture nonlinear features in the data, this technique has demonstrated superior results compared to linear PCA [5, 14, 15]. Here only the algorithmic application of kernel PCA will be given. For more thorough treatments, see [3, 13, 14, 15].

To form the statistical model of shape space $\mathbb{R}^n$, kernel PCA is performed as follows. Let $\{\mathbf{x}_1, ..., \mathbf{x}_N\} \subset \mathbb{R}^n$ be a set of aligned training shapes represented by binary masks where 1 is object and 0 is background and spread as vectors. These are samples from our shape space $\mathbb{R}^n$. Further, let the chosen kernel $k : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ representing the inner product in feature space be the radial basis function:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\tfrac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\rho^2}\right) \quad (4)$$

First, compute the $N \times N$ kernel matrix K with entries $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ taking training set elements pairwise. Second, perform the singular value decomposition on the centered kernel matrix $HKH = U\Sigma U^T$, with centering matrix $H = I - \frac{1}{N}\mathbf{1}\mathbf{1}^T$, identity matrix I of dimension $N \times N$, and $\mathbf{1} = [1\ 1...1]^T$ of dimension $N \times 1$. We now have matrix $U = [\mathbf{a}_1, ..., \mathbf{a}_N]$ containing the eigenvectors $\mathbf{a}_i = [a_{i1}, ..., a_{iN}]^T$ and diagonal matrix $\Sigma = \text{diag}(\lambda_1, ..., \lambda_N)$ capturing the corresponding eigenvalues. We now have a vector basis for the nonlinear feature space.

Thus, given a point $\mathbf{x} \in \mathbb{R}^n$ one can compute its projection $P\varphi(\mathbf{x}) \in \mathcal{H}$. For a given training set, the projection $P\varphi(\mathbf{x})$ is the shape most faithful to the training set. Since the map $\varphi : \mathbb{R}^n \rightarrow \mathcal{H}$ is not known, the pre-image $\hat{\mathbf{x}} \in \mathbb{R}^n$ of $P\varphi(\mathbf{x})$, that is $\hat{\mathbf{x}} = \varphi^{-1}(P\varphi(\mathbf{x}))$, cannot be easily computed. However, the method in [15] gives an algebraic expression to compute the approximate pre-image. This pre-image $\hat{\mathbf{x}}$ is the shape used to form the prior.

To obtain the pre-image $\hat{\mathbf{x}}$ of any aligned binary shape $\mathbf{x}$ we use the following expressions from [15] to form the pre-image $\hat{\mathbf{x}}$ given $\mathbf{x}$. Note that $\hat{\mathbf{x}}$ is a convex combination of the training set as judged by the nonlinear distance in feature space, $\|\varphi(\mathbf{x}_i) - P\varphi(\mathbf{x})\|^2$.

$$\mathbf{k}_x = [k(\mathbf{x}, \mathbf{x}_1), ..., k(\mathbf{x}, \mathbf{x}_N)]^T \quad \tilde{\mathbf{k}}_x = H(\mathbf{k}_x - \tfrac{1}{N}K\mathbf{1}) \quad (5)$$

$$\boldsymbol{\beta} = \left[\tfrac{1}{\sqrt{\lambda_1}}\mathbf{a}_1, ..., \tfrac{1}{\sqrt{\lambda_N}}\mathbf{a}_N\right]^T \tilde{\mathbf{k}}_x \qquad M = \sum_{i=1}^{N} \frac{1}{\lambda_i}\mathbf{a}_i\mathbf{a}_i^T \quad (6)$$

$$\boldsymbol{\gamma} = [\mathbf{a}_1, ..., \mathbf{a}_n]\boldsymbol{\beta} \qquad \tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + \tfrac{1}{N}(\mathbf{1} - \mathbf{1}^T\boldsymbol{\gamma}) \quad (7)$$

Finally, the pre-image is formed as

$$\hat{\mathbf{x}} = \frac{\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i(\frac{1}{2}(2 - \|\varphi(\mathbf{x}_i) - P\varphi(\mathbf{x})\|^2))\mathbf{x}_i}{\sum_{i=1}^{N} \tilde{\boldsymbol{\gamma}}_i(\frac{1}{2}(2 - \|\varphi(\mathbf{x}_i) - P\varphi(\mathbf{x})\|^2))} \quad (8)$$

using the distance in feature space

$$\|\varphi(\mathbf{x}_i) - P\varphi(\mathbf{x})\|^2 = (\tilde{\mathbf{k}}_x + 2H(\tfrac{1}{N}K\mathbf{1} - \mathbf{k}_{x_i}))^T M\tilde{\mathbf{k}}_x + \tfrac{1}{N^2}\mathbf{1}^T K\mathbf{1} + K_{ii} - \tfrac{2}{N}\mathbf{1}^T\mathbf{k}_{x_i} \quad (9)$$

5. PROPOSED ALGORITHM

Starting with a training set of binary shapes, we first define the kernel shape space by computing the matrices K, U, Σ, and M. To represent shapes, we use binary maps which have been shown to have higher fidelity shape projections compared to distance functions [5]. The following steps are cycled through repeatedly until convergence. The first iteration uses the user-initialization defining object and background while each subsequent iteration uses the previous segmentation.

First, compute object and background intensity histograms and smooth as necessary. Second, form the shape prior as the pre-image $\hat{\mathbf{x}}$ of the current segmentation $\mathbf{x}$ using expressions (5-9). Third, construct the graph using boundary weights as in (3), yet introducing a Bayesian prior into the regional term (2) as follows.

Our goal is to determine $P(\mathcal{O}|\mathcal{I})$ and Bayes rule tells us that $P(\mathcal{O}|\mathcal{I}) \propto P(\mathcal{I}|\mathcal{O})P(\mathcal{O})$. If we were to assume the prior to be uniform, $P(\mathcal{O}) = 1$, then its log-likelihood $\ln P(\mathcal{O}) = 0$, and it falls out of the expression as in (2). However, here we assume a non-uniform shape prior $P(\mathcal{O})$, and introducing a weight $0 \leq \mu \leq 1$ for relative shape influence, we have a new regional term:

$$\begin{aligned} R_p(\mathcal{O}) &= -(1-\mu)\ln P(\mathcal{I}_p|\mathcal{O}) - \mu\ \ln P(\mathcal{O}_p) \\ R_p(\mathcal{B}) &= -(1-\mu)\ln P(\mathcal{I}_p|\mathcal{B}) - \mu\ \ln P(\mathcal{B}_p) \end{aligned} \quad (10)$$

Since the pre-image has value between 0 and 1, we take the pre-image $\hat{\mathbf{x}}$ to directly be $P(\mathcal{O})$ and set $P(\mathcal{B}) = 1 - P(\mathcal{O})$. Finally, the min-cut of this graph yields a new segmentation that is used in the next iteration.

It is important to perform all shape computations with aligned shapes. Specifically, the training set should be aligned and pre-image formation should start with a aligned shape. For the purposes of this paper we assumed only rigid translation–no rotation or scaling. That said, before projecting we center the segmented shape according to its centroid, form the pre-image, and then undo the centering translation. More advanced registration may be necessary depending upon application [16, 17].

6. EXPERIMENTS

Segmentation was performed on three natural image sets and representative examples chosen to exhibit weak edges, clutter, and occlusion. Segmentation was performed on a Pentium IV 3.6 GHz processor with 2GB RAM and used the graph-cut software provided by [18]. All images are of size 120×160 and gray-scale. For all examples, kernel PCA training was performed in under a second, and segmentation–from initial segmentation through final iteration–also completed in less than a second. Each set of images uses its own training set of representative hand-segmented examples.

Parameters were fixed across all experiments with $\sigma = 20$ in the boundary term (3), ρ^2 in the kernel function (4) is the mean distance pairwise among training examples $\rho^2 = \langle \|\mathbf{x}_i - \mathbf{x}_j\|^2 / \|p - q\|^2 \rangle$, $\lambda = 1$ in (1) to equally weight the region and boundary terms, and $\mu = 0.4$ in (10) giving slight bias to intensity over shape. Neighborhoods of 8 were considered in graph construction. Convergence was reached in four or fewer iterations.

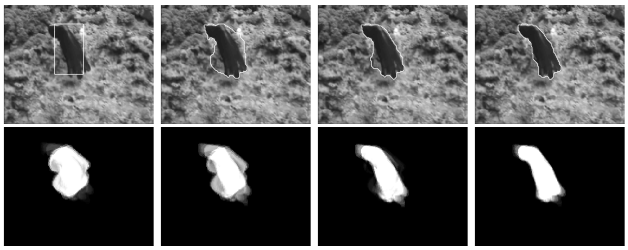

Fig. 2. Four iterations of octopus segmentation showing segmentation and shape prior pre-images.

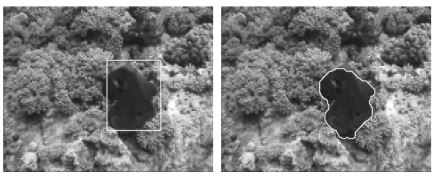

Fig. 3. Octopus of very different shape than Figures 1 and 2 to show variability of shape prior.

The first set of images involves an octopus with a training set of size 22. Figure 1 shows the octopus alongside a fish of similar intensity to demonstrate how the shape prior contains the segmentation from leaking into the fish. To see the iterative progress of segmentation and the shape prior, Figure 2 shows each at successive iterations. Figure 3 is shown to demonstrate the variation captured in the shape prior compared to Figure 2.

The second set of images involves a shark with fine sharp fins and a training set of size 49. Figure 4 shows the inability to capture the fins without shape and Figure 5 shows the ability to capture occluded regions.

The final set of images involves a couple holding hands with a training set of size 44, half the instances being of the man, half being of the woman. The couple is of similar intensity and so, without shape information, segmentation of one partner has opportunity to leak as demonstrated in Figure (6). This also demonstrates the specificity of combining shape and intensity to discriminate between figures based upon placement of the same initializing shape and underlying initial histograms.

7. CONCLUSION

In this note, we proposed how to incorporate a Bayesian shape prior into existing iterative graph cut segmentation methods. The shape model is learned from a set of training examples via kernel PCA and the shape prior generated by pre-image projection. Experiments on natural imagery showed the method to be accurate and efficient. Future work will explore segmenting multiple objects with the multi-label graph cut algorithm.

8. REFERENCES

[1] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," in *Int. Conf. on Computer Vision*, 2001, pp. 105–112.

[2] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D

Fig. 4. Necessity of shape to capture sharp fins: initial, without shape, with shape.

Fig. 5. Ability to capture occluded regions: initial, without shape, with shape.

Fig. 6. Couple of similar intensity showing leaking in absence of shape prior: initial, without shape, with shape.

image segmentation," *Int. J. of Computer Vision*, vol. 70, pp. 109–131, 2006.

[3] S. Mika, B. Scholkopf, A. Smola, M.Scholz K.R.Muller, and G.Ratsch, "Kernel PCA and de-noising in feature spaces," in *NIPS 11*, 1998.

[4] T. Cootes, C. Beeston, G. Edwards, and C. Taylor, "A unified framework for atlas matching using active appearance models," in *Proc. of Int. Conf. on Info. Proc. in Medical Imaging*, 1999, pp. 322–333.

[5] S. Dambreville, Y. Rathi, and A. Tannenbaum, "Shape-based approach to robust image segmentation using kernel PCA," in *Computer Vision and Pattern Recognition*, 2006, pp. 17–22.

[6] P. Das, O. Veksler, V. Zavadsky, and Y. Boykov, "Semiautomatic segmentation with compact shape prior," in *Canadian Conf. on Computer and Robot Vision*, 2006, pp. 28–36.

[7] G. Slabaugh and G. Unal, "Graph cuts segmentation using an elliptical shape prior," in *Int. Conf. on Image Processing*, 2005, pp. 1222–5.

[8] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," in *ACM Trans. on Graphics (SIGGRAPH)*, 2004.

[9] D. Freedman and T. Zhang, "Interactive graph cut based segmentation with shape priors," in *Computer Vision and Pattern Recognition*, 2005, pp. 755–762.

[10] M. Kumar, P. Torr, and A. Zisserman, "ObjCut," in *Computer Vision and Pattern Recognition*, 2005, pp. 18–23.

[11] V. Kolmogorov and Y. Boykov, "What metrics can be approximated by geo-cuts, or global optimization of length/area and flux," in *Int. Conf. on Computer Vision*, 2005, pp. 564–571.

[12] N. Xu, R. Bansal, and N. Ahuja, "Object segmentation using graph cuts based active contours," in *Computer Vision and Pattern Recognition*, 2003, pp. 46–53.

[13] B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," Tech. Rep., Max-Planck-Institute fur biologische Kybernetik, 1996.

[14] D. Cremers, T. Kohlberger, and C. Schnorr, "Shape statistics in kernel space for variational image segmentation," *Pattern Recognition*, vol. 36, pp. 1929–1943, 2003.

[15] Y. Rathi, S. Dambreville, and A. Tannenbaum, "Statistical shape analysis using kernel PCA," in *IS&T/SPIE Symposium on Electronic Imaging*, 2006.

[16] Y. Keller and A. Averbuch, "Fast gradient methods based on global motion estimation for video compression," *Trans. on Circuits and Systems for Video Technology*, vol. 13, no. 4, 2003.

[17] A. Tsai, A. Yezzi, W. 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," *Trans. on Medical Imaging*, vol. 22, no. 2, pp. 137–154, 2003.

[18] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.