USER DRIVEN SPARSE POINT-BASED IMAGE SEGMENTATION

Sachin Meena[⋆] *Kannappan Palaniappan*[⋆] *Guna Seetharaman*[†]

[⋆]Department of Computer Science, University of Missouri-Columbia, MO 65211 USA
[†]US Naval Research Laboratory, Washington, DC 20375 USA

ABSTRACT

Reducing the amount of user driven input for interactive image segmentation enables faster and more precise foreground extraction of objects. A sparse collection of labeled seed points sampled over image regions can be quickly provided by the user using a few mouse clicks. Seed points are used for training an Elastic Body Spline classifier mapping function. We evaluate the efficiency and accuracy of user defined point inputs compared to fully manual boundary drawing that can be time consuming and automatic image segmentation methods that may not have sufficient accuracy. We show that using an average of just 8 labeled pixels (*i.e.* sparse set of seed points) the proposed EBS foreground-background thresholding method can achieve 90 percent accuracy compared to manual ground truth on the Berkeley BSDS500 benchmark.

Index Terms— Interactive image segmentation, seed points, semi-supervised regression, elastic body splines.

1. INTRODUCTION

User driven image editing and segmentation tasks often are based on lines (scribbles or strokes) [1], contours [2], [3] and bounding rectangles [4]. Medical, scientific and defense applications often require high accuracy [5–7] whereas consumer applications focus more on visual aesthetics and quality [2], [4]. These different requirements effect the degree of precision and consequently time required to provide the various types of user driven inputs. A sparse set of labeled seed points was recently proposed in [8], [9], [10] as a novel approach to reduce the user workload. Points are defined by just single user clicks, do not require drawing lines, contours or rectangles and can be rapidly moved, deleted or corrected to interactively improve segmentation performance. We have developed a region centric point-based interactive segmentation approach that requires less user interaction compared to the BeyondDots method which defines points sampled around the foreground object boundary [10]. Our proposed interactive process for point-based image segmentation is shown in Fig. 1 and a sample result using just four points on a basilica image from the BSDS500 [11] image segmentation benchmark dataset is shown in Fig. 2.

Earlier work in this area are live-wire [2] and intelligent scissors [3] users provide some information about the location of the object boundary which is then used to guide the image segmentation task. Both live-wire [2] and intelligent scissors [3], are edge-based methods and highly susceptible to boundary noise and can leak through weak object boundaries. In active contour models user initialize a curve close to the boundary of desire object and then the curve is evolved based on local edge information [12] or region information [13]. One of the main drawbacks of these methods is solutions are only locally optimum hence the final results are greatly affected by the initial input provided by the user. Graph-based methods such as [1], [4] and [14] frame the interactive segmentation task as a *min-cut* problem, which is solved by the max flow algorithm. It finds the minimum cost cut that best separates the foreground from the background. Random Walker [15] is another graph based algorithm which finds the probability of reaching a foreground or background pixel from a unlabeled pixel. Here a random walker starts from a unlabeled pixel and then jumps from pixel to pixel until it reaches a foreground or a background pixel. More recent work done in the area of interactive image segmentation include [16], [17] and [18].

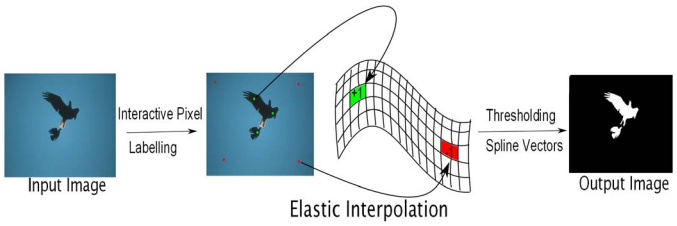


Fig. 1. Interactive image segmentation process showing user driven labeled seed selection (foreground seeds in green color, background labels in red color), elastic body spline manifold mapping, followed by segmented binary output image.

In this work we use Elastic Body Splines (EBS) which are analytic solutions of the Navier equilibrium PDE that models the deformation of an elastic body. EBS have been applied to the task of biomedical image registration [19] and in this work we propose to use it for interactive image segmentation. In Section 2 we provide a brief description of the original EBS based image registration problem and adapt the EBS framework for interactive image segmentation and include relevant

(a) 4 marked seeds (b) GT (c) Proposed method F-score:0.9958

Fig. 2. Sample image segmentation of a building (image from BSD500) using just four input seed points shown in green (foreground) and red (background). Detailed architectural structures are accurately segmented. From left to right Seed Points, Ground Truth, EBS (proposed method).

mathematical details. In Section 4 we provide quantitative and qualitative results of the proposed framework and finally in section 5 we provide a summary of our work along with possible directions for future work.

2. EBS FRAMEWORK FOR INTERACTIVE IMAGE SEGMENTATION

Elastic Body Splines (EBS) belong to the class of 3D splines suggested by Davis in [19] and have been mainly used for elastic registration of biomedical images. In [9] the authors proposed to use interactive point inputs with EBS interpolation for image foreground segmentation. Our classifier spline function,

$$\mathbf{G}(\vec{x}) = \beta r(\vec{x})\mathbf{I} - \vec{x}\vec{x}^T/r(\vec{x}) \tag{1}$$

is obtained as the solution to the Navier PDE,

$$\mu\nabla^2\vec{u}(\vec{x}) + (\mu+\lambda)\nabla[\nabla\cdot\vec{u}(\vec{x})] = \vec{f}(\vec{x}) \tag{2}$$

where $\beta = 8(1-\nu)-1$ and $\nu = \frac{\lambda}{2(\lambda+\mu)}$. $\vec{u}(\vec{x})$ is the displacement, $\vec{f}(\vec{x})$ the force field, μ and λ are the Lame coefficients, ∇^2 and ∇ denote the Laplacian and gradient respectively. The force function,

$$\vec{f}(\vec{x}) = \vec{c}/r(\vec{x}) \tag{3}$$

decreases symmetrically with increasing radial distance $r(\vec{x})$. For more details on EBS we refer [9], [8].

Gaussian elastic body splines (GEBS) belong to the class of elastic splines suggested by Kohlrausch in [20] for biomedical image registration. The spline function,

$$\begin{aligned}\mathbf{G}(\vec{x}) = \Bigg[\Bigg(&(4(1-\nu)-1)\frac{\mathrm{Erf}(\hat{r})}{r} \\ &- \sqrt{\frac{2}{\pi}}\sigma\frac{e^{-\hat{r^2}}}{r^2} + \sigma^2\frac{\mathrm{Erf}(\hat{r})}{r^3}\Bigg)\mathbf{I} \\ &+ \left(\frac{\mathrm{Erf}(\hat{r})}{r^3} + 3\sqrt{\frac{2}{\pi}}\sigma\frac{e^{-\hat{r^2}}}{r^4} - 3\sigma^2\frac{\mathrm{Erf}(\hat{r})}{r^5}\right)\vec{x}\vec{x}^T\Bigg].\end{aligned} \tag{4}$$

is solution to (2) by using the force function,

$$\vec{f}(\vec{x}) = \vec{c_i}\frac{1}{(\sqrt{2\pi}\sigma)^3}exp\frac{-r^2}{2\sigma^2} \tag{5}$$

where

$$\mathrm{Erf}(x) = \frac{2}{\pi}\int_0^x e^{-t^2}dt \tag{6}$$

We first demonstrated using the spline (4) for interactive image segmentation [8].

For an interactive image segmentation task we learn the classification function $\vec{d}(\vec{x})$ as the following spline form,

$$\vec{d}(\vec{x}) = \mathbf{A}\vec{x} + \vec{b} + \sum_{i=0}^{N}\mathbf{G}_{EBS}(\vec{x}-\vec{p_i})\vec{c_i}. \tag{7}$$

where $\mathbf{G}_{EBS}$ is given by (1) such that,

$$\sum_{i=0}^{N}\vec{c_i} = \vec{0} \quad , \sum_{i=0}^{N}\vec{c_i}p_{ij} = \vec{0} \quad j = 1,\ldots,5 \tag{8}$$

where $\vec{x}$ is a feature vector, $\mathbf{G}$ is a spline matrix, $\vec{c_i}$ are the spline coefficients, $\vec{p_i}$ are the labeled pixels and the matrix $\mathbf{A}$ and vector b are coefficients for the linear mapping. The constraints in (8) ensure that the system of equations (7) has a unique solution.

The supervised classifier function $\vec{d}(\vec{x})$ is learned from user supplied labeled seed points such that $\vec{d}(\vec{x})) = +\vec{1}^T$ for the foreground pixels and $\vec{d}(\vec{x}) = -\vec{1}^T$ for the background pixels. Let $\vec{w}$ be the vector of all the EBS coefficients given as,

$$\vec{w} = \begin{bmatrix}\mathbf{C}_F & \mathbf{C}_B & \mathbf{A} & \vec{b}^T\end{bmatrix}^T \tag{9}$$

where $\mathbf{C}_F$ and $\mathbf{C}_B$ are the elastic coefficients corresponding to the foreground and the background seed pixels.

$$\mathbf{C}_F = \left[\vec{c}_1^T \ldots \vec{c}_f^T\right], \quad \mathbf{C}_B = \left[\vec{c}_1^T \ldots \vec{c}_b^T\right].$$

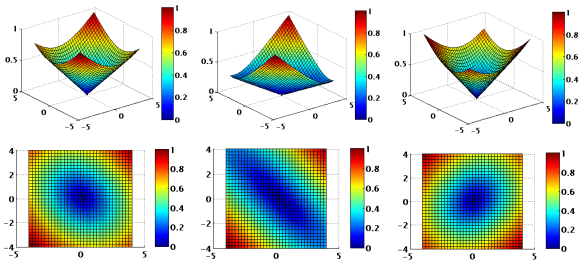
The EBS mapping determined by the weights $\vec{w}$ that we are solving for, is given by the relationship, $\mathbf{Y} = \mathbf{L}\vec{w}$, where $\mathbf{Y}$ is the set of displacement vectors for the user defined seed points. The EBS coefficients are estimated by solving the matrix equation,

$$\vec{w} = \mathbf{L}^{-1}\vec{Y} \tag{10}$$

where,

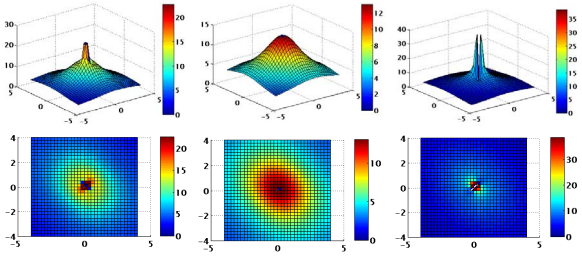
$$\mathbf{L} = \begin{bmatrix}\mathbf{K} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{O}\end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix}\mathbf{G}_{FF} & \mathbf{G}_{FB} \\ \mathbf{G}_{BF} & \mathbf{G}_{BB}\end{bmatrix} \tag{11}$$

$$\mathbf{G}_{FF}(\vec{r}) = \begin{bmatrix}\mathbf{G}_{11}(\vec{r}_{11}) & \ldots & \mathbf{G}_{1f}(\vec{r}_{1f}) \\ \vdots & & \vdots \\ \mathbf{G}_{f1}(\vec{r}_{11}) & \ldots & \mathbf{G}_{ff}(\vec{r}_{ff})\end{bmatrix} \tag{12}$$



(a) $\nu = 0.3$ (b) $\nu = 0.7$ (c) $\nu = 1.4$

Fig. 3. EBS classification functions for different μ values.



(a) $\sigma = 0.8$ (b) $\sigma = 1.0$ (c) $\sigma = 1.2$

Fig. 4. GEBS Eq. (4) classification functions for different σ values.

with $\vec{r}_{ij}$ is the distance between feature vectors $\vec{p}_i$ and $\vec{p}_j$ of the foreground seed points. Hence, $\mathbf{G}_{FF}$ is the matrix of EBS (1) functions defined only over the foreground pixels. $\mathbf{G}_{FB}$, $\mathbf{G}_{BB}$ and $\mathbf{G}_{BF}$ are similarly defined,

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_F & \mathbf{I}_F \\ \mathbf{P}_B & \mathbf{I}_B \end{bmatrix}, \tag{13}$$

where the $\mathbf{I}_F$ and $\mathbf{I}_B$ are sets of identity matrices,

$$\mathbf{I}_F = \begin{bmatrix} \mathbf{I}_1 \dots \mathbf{I}_f \end{bmatrix}, \quad \mathbf{I}_B = \begin{bmatrix} \mathbf{I}_1 \dots \mathbf{I}_b \end{bmatrix}, \tag{14}$$

$$\mathbf{P}_F = \begin{bmatrix} x_{11}\mathbf{I} & \dots & x_{15}\mathbf{I} \\ \vdots & & \vdots \\ x_{f1}\mathbf{I} & \dots & x_{f5}\mathbf{I} \end{bmatrix} \tag{15}$$

where x_{ij} is the j^{th} feature for i^{th} pixel. $\mathbf{P}_B$ is similarly defined. The vector $\vec{Y}$ consists of $\vec{Y}_F$ with values +1 for the foreground seed points and $\vec{Y}_B$ with values -1 for the background seed points,

$$\vec{Y} = \begin{bmatrix} \vec{Y}_F & \vec{Y}_B & \vec{O} \end{bmatrix}^T \tag{16}$$

where $\vec{O}$ is a vector of zeros. Once the classification function (4) is learned we use zero (midpoint between -1 and $+1$) to threshold the vector $\vec{d}(\vec{x})$ and assign label $\ell(\vec{x})$ to a pixel by taking consensus among vector elements of $\vec{d}(\vec{x})$,

$$\ell(\vec{x}) = \begin{cases} foreground, & \text{if majority } \vec{d}(\vec{x}) \text{ elements} \geq 0 \\ background, & \text{if majority } \vec{d}(\vec{x}) \text{ elements} < 0. \end{cases} \tag{17}$$

Parameter Selection The spine functions EBS and GEBS have parameters that need to be tuned for the interactive image segmentation task. In (1) the EBS Poisson parameter ν is related to the elastic property of the material. GEBS (4) has two parameters to tune namely ν Poisson's ratio and σ the width of the Gaussian force function. The shape of the EBS function for $x = (x_1, x_2)$ the two spatial features is illustrated in Fig. 3 for three values of ν. We used $\nu = 0.3$ which performed best for EBS. Similarly, for GEBS (4) we used $\sigma = 1.0$ with $\nu = 0.3$ and the shape of several classification splines are shown in Fig. 4.

No.	Methods	Grab-Cut [9]	BSDS500
1	EBS	**0.8770±0.0993**	**0.9009±0.0594**
2	GEBS	0.7370±0.2273	0.7852±0.2163
3	RW	0.6928±0.2403	0.6189±0.2483
4	GC	0.3869±0.3485	0.4786±0.3188

Table 1. Quantitative Evaluation : Average F-measure for 50 images from Grab-cut data set [4] using an average of 8.46 marked pixels per image. Average F-measure for 50 images from Berkeley data set [11] using an average of 8.12 marked pixels per image.

3. EXPERIMENTAL RESULTS

In our experiments we have used images from the BSDS500 [11] data set. We have randomly selected 50 images from the data set and then performed the user driven image segmentation using four methods, Elastic Body Spline (EBS), Gaussian Elastic Body Spline (GEBS) [8], Random Walker (RW) [15] and Graph Cut (GC) [1]. Ten or fewer pixels are marked as foreground/background and the same initialization points are used for all the four methods. In [21] the authors have provided the performance accuracies for different automatic methods using the BSDS500 dataset. The best result has an accuracy of 0.71 while the accuracy for human agreement is .79 [21]. As we can see from the Table I EBS achieves an F-score of 0.90 using just 8 marked seed points per image. Table I shows that EBS is about 12 percent better than GEBS, while outperforming RW and GC by 28 and 42 percent respectively.

4. CONCLUSIONS

The choice of the classifier spline function is critical to good interpolation performance for user driven image segmentation with the EBS radial force function outperforming GEBS

(a) 10 marked pixels (b) 0.8955 (c) 0.5749 (d) 0.7705 (e) 0.1085

(f) 8 marked pixels (g) 0.8861 (h) 0.8196 (i) 0.3986 (j) 0.5408

(k) 10 marked pixels (l) 0.9041 (m) 0.8520 (n) 0.3823 (o) 0.0767

(p) 10 marked pixels (q) GT (r) EBS 0.9584 (s) GEBS 0.7851 (t) RW 0.6941 (u) GC 0.6634

Fig. 5. Quantitative results using images from Berkeley database [11] with F-scores shown below each result. From left to right input seeds superimposed on the original image, ground truth mask, EBS, GEBS [8], Random Walker [15] (RW), Graph-cut [1] (GC). No pre-processing or post processing is performed for the results shown.

which uses a Gaussian force function. EBS also consistently outperformed Random Walk and Graph-Cut methods on a subset of the BSDS500 image segmentation benchmark dataset by a significant margin, confirming the previous results using the Grab-Cut data set [9]. We have shown that using an average of just 8 labeled pixels or sparse set seed points the proposed EBS foreground-background thresholding method can achieve 90 percent accuracy compared to the consensus manual segmentation. This compares favorably with an F-measure for human agreement of 79 percent and best automatic algorithm accuracy of 71 percent.

Acknowledgments

This research was partially supported by U.S. Air Force Research Laboratory (AFRL) under agreement AFRL FA875014-2-0072. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of AFRL or the U.S. Government.

5. REFERENCES

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

[2] A. X. Falcao and J. K. Udupa, "A 3D generalization of user-steered live-wire segmentation," *Medical Image Analysis*, vol. 4, no. 4, pp. 389–402, 2000.

[3] E. N. Mortensen and W. A. Barrett, "Intelligent scissors for image composition," in *Proc. Conf. on Computer Graphics and Interactive Techniques*, 1995, pp. 191–198.

[4] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, pp. 309–314, 2004.

[5] F. Bunyak, A. Hafiane, and K. Palaniappan, "Histopathology tissue segmentation by combining

fuzzy clustering with multiphase vector level sets," in *Software Tools and Algorithms for Biological Systems*. Springer, 2011, pp. 413–424.

[6] S. K. Nath, F. Bunyak, and K. Palaniappan, "Robust tracking of migrating cells using four-color level set segmentation," *Lecture Notes in Computer Science (ACIVS)*, vol. 4179, pp. 920–932, 2006.

[7] K. Palaniappan, R. Rao, and G. Seetharaman, "Wide-area persistent airborne video: Architecture and challenges," in *Distributed Video Sensor Networks: Research Challenges and Future Directions*. Springer, 2011, pp. 349–371.

[8] S. Meena, V. B. S. Prasath, K. Palaniappan, and G. Seetharaman, "Elastic body spline based image segmentation," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2014, pp. 4378–4382.

[9] S. Meena, K. Palaniappan, and G. Seetharaman, "Interactive image segmentation using elastic interpolation," in *IEEE Int. Symposium on Multimedia (ISM)*, 2015, pp. 307–310.

[10] T. Windheuser, T. Schoenemann, and D. Cremers, "Beyond connecting the dots: A polynomial-time algorithm for segmentation and boundary estimation with imprecise user input," in *IEEE Conf. Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 717–722.

[11] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *IEEE Int. Conf. Computer Vision*, 2001, pp. 416–423.

[12] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, pp. 321–331, 1988.

[13] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. on Image Processing*, pp. 266–277, 2001.

[14] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive image segmentation using an adaptive gmmrf model," in *European Conf. on Computer Vision*, 2004, pp. 428–441.

[15] L. Grady, "Random walks for image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, no. 11, pp. 1768–1783, 2006.

[16] J. Xu, M. D. Collins, and V. Singh, "Incorporating user interaction and topological constraints within contour completion via discrete calculus," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2013, pp. 1886–1893.

[17] C. Nieuwenhuis and D. Cremers, "Spatially varying color distributions for interactive multilabel segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 1234–1247, 2013.

[18] J. Wu, Y. Zhao, J.-Y. Zhu, S. Luo, and Z. Tu, "Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2014, pp. 256–263.

[19] M. H. Davis, A. Khotanzad, D. P. Flamig, and S. E. Harms, "A physics-based coordinate transformation for 3-d image matching," *IEEE Trans. on Medical Imaging*, pp. 317–328, 1997.

[20] J. Kohlrausch, K. Rohr, and H. Siegfried Stiehl, "A new class of elastic body splines for nonrigid registration of medical images," *Journal of Mathematical Imaging and Vision*, pp. 253–280, 2005.

[21] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.