

# Segmentation from a Box

Leo Grady                      Marie-Pierre Jolly

Siemens Corporate Research — Image Analytics and Informatics  
755 College Rd., Princeton NJ

Leo.Grady (Marie-Pierre.Jolly)@siemens.com

Aaron Seitz

University of California Riverside — Department of Psychology  
900 University Ave., Riverside CA

aseitz@ucr.edu

## Abstract

*Drawing a box around an intended segmentation target has become both a popular user interface and a common output for learning-driven detection algorithms. Despite the ubiquity of using a box to define a segmentation target, it is unclear in the literature whether a box is sufficient to define a unique segmentation or whether segmentation from a box is ill-posed without higher-level (semantic) knowledge of the intended target. We examine this issue by conducting a study of 14 subjects who are asked to segment a boxed target in a set of 50 real images for which they have no semantic attachment. We find that the subjects do indeed perceive and trace almost the same segmentations as each other, despite the inhomogeneity of the image intensities, irregular shapes of the segmentation targets and weakness of the target boundaries. Since the subjects produce the same segmentation, we conclude that the problem is well-posed and then provide a new segmentation algorithm from a box which achieves results close to the perceived target.*

## 1. Introduction

Image segmentation is generally considered to be an ill-posed problem, so much so that it is even difficult to agree in the literature how to measure the similarity of two hand-drawn image segmentations created by different people (see [10, 20, 7, 31] for different approaches to this problem). In contrast, the segmentation of a single, semantically-defined object is typically considered to be well-enough defined that algorithm performance can be measured against ground truth and compared. Segmentation tasks of this type feature regularly in medical imaging, for which there exists an annual “Segmentation Grand Challenge” offered for various medical targets (e.g., [11]).

In determining why the unconstrained image segmentation task is ill-posed and the targeted image segmentation task is (relatively) well-posed, it is important to recognize several differences between these problems. Specifically, the algorithms designed to meet the unconstrained image segmentation challenge (e.g., [26, 9, 14, 2]) and the targeted image segmentation task (e.g., [11]) have different properties:

1. The targeted segmentation algorithm returns a single object, while the unconstrained image segmentation algorithm may return an unknown number of objects depending on the level of fine detail considered (e.g., whether a “tire” is a separate object from a “car”).
2. The objective of the targeted segmentation algorithm is to find a semantically-defined object while a valid output of the unconstrained segmentation algorithm could be homogeneous regions or textures.
3. Some prior knowledge of the intended target object is built into a targeted image segmentation algorithm, while an unconstrained segmentation algorithm may be more general-purpose.

It has not been well-studied in the literature which of these features are sufficient to constrain the segmentation problem to be well-posed.

The interactive segmentation algorithms which have been recently studied (e.g., [6, 1, 12, 4, 30]) have more in common with the class of targeted segmentation algorithms in the sense that the objective is typically to segment a single object, the object is semantically defined (as perceived by the user) and some prior knowledge of the intended target is known (provided by the user interaction). In this sense, the interactive segmentation problem may be considered as relatively well-posed and the performance of an interactive al-

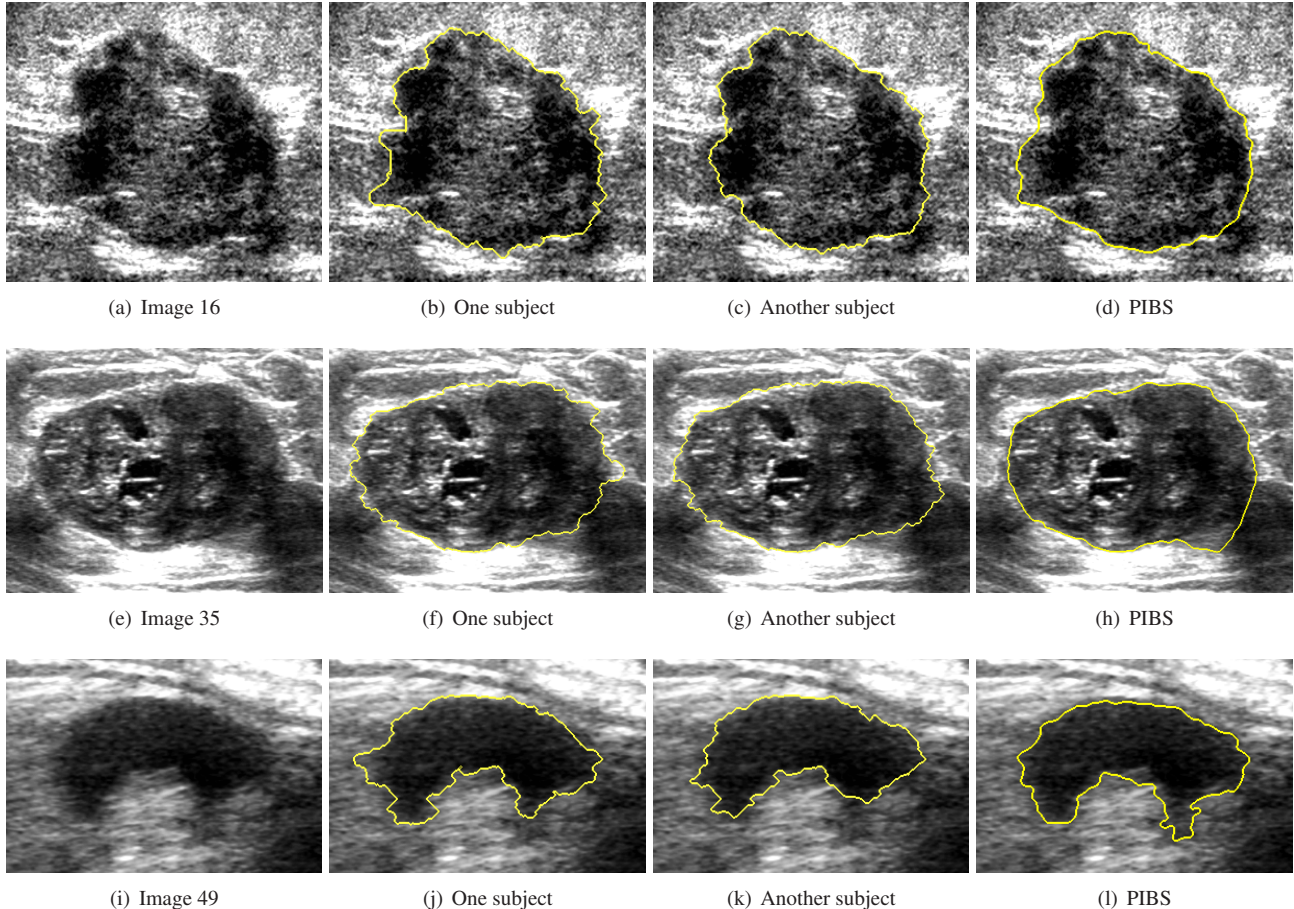


Figure 1. Example images and segmentations from our study. Despite the lack of semantic understanding of these images, weak boundaries, object/background internal inhomogeneity and nonstandard shapes, our subjects produced remarkably similar segmentations. The algorithm proposed here (PIBS) is capable of capturing some of these difficult cases to provide segmentations very similar to those perceived by the subjects.

gorithm can be quantitatively evaluated (as attested by several studies [19] and interactive databases [23, 24]). Since the class of interactive segmentation algorithms is relatively well-posed, they are an attractive subject of study since they are also intended to be general-purpose (e.g., appearing in commercial products which allow a user to segment any object of interest). This general-purpose nature means that a strong solution to the interactive segmentation problem can be easily repurposed to create an automatic targeted segmentation algorithm simply by simulating the user interactions to direct the algorithm to the intended target (e.g., [34, 15]). The “box” interface of [23] is particularly attractive for this purpose, since many of the recent object detection algorithms [33, 29] output a box around the intended target. If this automatically-generated box could be passed as an input to a high-performance, general-purpose segmentation algorithm then a straightforward pipeline exists for creating an automatic segmentation algorithm for any feature-driven target.

The model of interactive segmentation algorithms for general-purpose targeted image segmentation does create some concerns. For example, it is known that there is some effect on performance quality which is created by applying different interaction strategies [28]. Furthermore, interactive methods have taken several different kinds of inputs from a user, from seeds/scribbles [6] to approximate initial segmentations such as an ellipse [25] or a box [23, 18, 17] and there is no generally accepted guidance for how a user should drive these inputs. Specific to the box interface, it is unclear how large the box should be drawn relative to the intended target and whether or not the specification of a box around a segmentation is sufficient to define a general segmentation target.

Before designing a general-purpose algorithm to produce a segmentation from a box, a major question is whether a box is sufficient to define a unique target object. Within the segmentation literature, some authors have offered the perspective that these objects must be defined

semantically [3], while others have preferred to describe segmentations in terms of an optimization of the region which is described by low-level features (e.g., which is homogeneous, smooth or possessing a regularized boundary [16, 22, 32]). Therefore, it is unclear whether a general-purpose box-driven segmentation algorithm must include some semantic understanding of the image to match the intent of a user or whether a low-level description is sufficient. Due to the large number of possible semantic objects, a general-purpose segmentation algorithm of this sort would be challenging to design if semantic information were necessary to define a segmentation target from a box.

In this paper we attempt to understand whether it is possible to create a general-purpose segmentation algorithm from a box input in the absence of semantic information about the segmentation target. Specifically, our initial purpose was to determine whether a large number of people produced the same segmentation, given a box, even when the subjects did not possess a semantic understanding of the image. Rather than creating an artificial set of images without semantic content, we instead used cropped natural images obtained from an ultrasound machine. Ultrasound images are known for being noisy and difficult for trained experts to interpret, even when they have access to the entire image. In this case, none of our subjects had access to more than a small portion of the image and none of our subjects had any background in reading ultrasound images (although some, but not all, had training in image processing). Despite these factors, the segmentations were remarkably consistent across subjects and across images, even for those images which had very little homogeneity, poor boundary definition or for which the identified segmentation had a irregular shape. Figure 1 shows several images in our study with the segmentations obtained by some of the subjects.

Having established that there is a unique answer for the segmentation box problem in this dataset, the next question was whether an algorithm could be created to achieve the target segmentations. We define a low-level segmentation algorithm (PIBS) as a joint functional of a nonparametric intensity model, boundary consistency and a boundary regularization which may be optimized iteratively. This PIBS algorithm is able to achieve strong performance over a range of the target images in this set, although some of the images continue to pose challenges.

Due to the unconventional nature of this paper, we have disrupted the usual format of describing method and then results. First, we detail the study design, analysis method and results in Section 2. In this section we establish the consistency of box-driven segmentations across images and subjects and demonstrate that the results are neither random nor trivial. Having used the study section to establish that there is an agreed target for a segmentation algorithm, we then detail a new segmentation algorithm (PIBS) which is

designed to match the target box-driven segmentations. The PIBS algorithm is developed and evaluated in Section 3. Final conclusions are drawn in Section 4.

## 2. Segmentation consistency from a box

The main purpose of our study was to determine whether or not the segmentation box problem was tractable by determining if different people perceived the same target object in a series of natural, noisy, but semantic-free, images. Specifically, we say that the problem is tractable if the specification of a box was sufficient to elicit a unique segmentation across a wide range of people. In other words, if everybody agrees on the segmentation implied by a box then the algorithmic goal becomes the design of an algorithm which achieves the target segmentation. However, if every person perceives a different segmentation, then no single algorithm can possibly achieve the correct segmentation, i.e., the box problem would be *ill-posed*. We first describe the study design before detailing the analysis and interpretation of the results.

### 2.1. Study design

Fifty test images were used in this study which were obtained from breast and abdominal, scan-converted, monochromatic, B-mode and elastography ultrasound acquisitions. In each of these images, the authors identified some segmentation target “object” within the images and drew a rectangular, axis-aligned, box around the target such that the author-perceived target was completely contained inside the box, but the box was not significantly larger than necessary to completely contain the target. The authors have no training in reading ultrasound images and therefore were uninformed about the anatomy being displayed (other than knowing that the images were obtained from breast and abdominal scans), meaning that the targets identified by the authors may or may not have clinical relevance. However, clinical relevance was not the concern of our study, but instead our goal was to determine the consistency of object perception from a box across individuals. The authors chose target objects to box using a variety of criteria (and clarity of the object perception) to avoid any bias in the selection toward a particular object description (e.g., homogeneity, shape, etc.). Once a box was defined by the authors on an image, the portion of the image outside the box was removed to exclude any contextual information.

The authors created a tool which proceeded in a consistent order through all 50 images and allowed subjects to draw a segmentation which was saved before proceeding to the next image. The tool presented only the portion of the image which was contained in the box drawn by the authors. In order to control for a varying degree of effort made by subjects in producing fine-detail segmentations, the authors provided them with the intelligent scissors/live

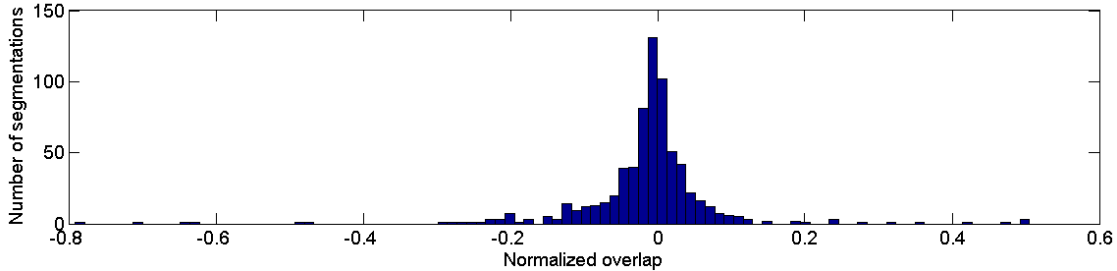


Figure 2. The histogram of subject-drawn segmentations relative to the mean segmentation for each image. Negative values indicate undersegmentation relative to the mean and positive values indicate oversegmentation relative to the mean. A Kolmogorov-Smirnov test on each image suggests that the responses are normally distributed ( $p = 6.7e^{-11}$ ).

wire tool [21, 8] which allowed them to click control points on the boundary such that these control points were connected automatically using the tool. The users could override the connection made by intelligent scissors by pressing and holding the mouse button to perform a freehand draw.

Fourteen subjects were enlisted in this study. None of the subjects had any medical training that would provide a semantic understanding of the (context-free) objects presented to them. Some of the subjects had training in image processing and medical imaging, but other subjects had training in neither. The purpose of the experiment was not revealed to any subject. Before beginning the task, the authors explained that the images the subjects were about to see were cropped out of a larger image and that their goal was to use the tool to segment an object they perceived within the image that satisfied the following criteria:

1. Each image contained only one target object.
2. The object was completely contained inside the image (box), i.e., the object did not touch the sides of the image (box).
3. The box was selected from the larger image such that the perceived object “roughly filled” the box (i.e., the “tightness prior” of [18]).

These instructions were read from a script to avoid individually biasing any of the subjects. The task required approximately 20 minutes to complete. Our study implicitly assumed that the target object was genus zero. Although this criterion was not communicated to the subjects, there were no situations in which the subjects produced target segmentations of higher genus.

## 2.2. Data analysis

Our purpose in designing this study was to see if the same segmentation target was perceived within a box by a majority of people. However, even using the same tool, we cannot expect each person to precisely identify the same segmentation even if they appear to perceive the same target. Consequently, we must formally define hypotheses

which can be used to quantitatively evaluate our data. We consider two hypotheses: 1) The segmentations are normally distributed about a mean segmentation, 2) The segmentations are different than the segmentations obtained by using a control set of trivial segmentations. Finally, we apply the Probabilistic Rand Index (PRI) from the literature [31], which is the standard measure of agreement between a test segmentation and a set of segmentations.

### 2.2.1 Segmentation distribution

Here we consider the hypothesis: The segmentations are normally distributed about a mean segmentation for each image. For each of  $K = 50$  images and  $S = 14$  subjects, a binary segmentation mask  $M_k^s$  was used to represent the segmentation. A probabilistic segmentation image  $R_k$  was computed such that each pixel  $i$  represented the probability that the pixel was included in some segmentation by computing

$$R_k(i) = \frac{1}{S} \sum_{s=0}^S M_k^s(i). \quad (1)$$

The mean segmentation image  $G_k$  was computed from  $R_k$  via

$$G_k(i) = \begin{cases} 1 & \text{if } R_k(i) > \frac{1}{2}, \\ 0 & \text{else.} \end{cases} \quad (2)$$

For each image and subject, a normalized overlap score  $\omega_k^s$  was calculated via

$$\omega_k^s = \frac{\sum_i M_k^s(i) - G_k^s(i)}{\sum_i G_k^s(i)}. \quad (3)$$

Note that  $\omega_k^s$  may be positive or negative. These normalized overlap scores are displayed for each subject and image in Figure 2.

To test for a normal distribution, we ran a Kolmogorov-Smirnov test on  $\omega_k^s$  to show that the segmentations for each image are normally distributed ( $p = 6.7185e^{-11}$ ) and the averaged response for each subject was also normally distributed ( $p = 0.0013$ ). A T-test may also be used to determine if the distribution of errors for an image is significantly

different from zero. The results of the T-test are shown in Figure 3, which show that only in one image is the distribution of errors significantly different than zero. Hartigan’s dip test shows that, except for image 27, there is no evidence of a bimodal distribution (at 0.05 significance) for the  $\omega$  values.

### 2.2.2 Control segmentations

For each image, two types of control segmentation were generated. The first control was a purely random segmentation in which each pixel was randomly assigned a value in the interval  $[0,1]$  (from a uniform distribution) and a cut-off was established such that the number of pixels above threshold were equal to  $\sum_i G_k(i)$ . The second control segmentation was generated by creating an ellipse centered in the center of the box, with an aspect ratio equal to the aspect ratio of the box and an area equal to a third of the total image size.

### 2.2.3 Probabilistic Rand Index

A standard measure in the literature for comparing a test segmentation to another segmentation is the Rand Index, which has unity value when the segmentations are exactly the same. However, to compare a test segmentation to a set of “ground truth” segmentations which do not completely agree, the Probabilistic Rand Index was defined [31]. Note that we did not employ the *normalized* PRI defined in [31] since this measure depends on the ability to define an expected PRI which requires some assumptions about the expected distribution of segmentations. However, since the purpose of our study was to measure the consistency between segmentations in this situation, the normalized PRI is inappropriate.

The Probabilistic Rand Index for a test segmentation  $T_k$ , compared against a set of “ground truth” segmentations (represented with  $M_k^s$ ) was computed as

$$\text{PRI}(T_k) = \frac{1}{\binom{N}{2}} \sum_{i,j,i < j} p(i,j)c(i,j) + (1 - p(i,j))(1 - c(i,j)), \quad (4)$$

where  $N$  is the number of pixels in the image,  $p(i,j)$  is defined by

$$p(i,j) = \frac{1}{S} \sum_s \mathbf{I}(M_k^s(i) = M_k^s(j)), \quad (5)$$

$c(i,j)$  is defined by

$$c(i,j) = \mathbf{I}(T_k(i) = M_k^s(j)), \quad (6)$$

and

$$\mathbf{I}(a = b) = \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{else.} \end{cases} \quad (7)$$

The PRI was computed for each segmentation produced by each subject against the set of segmentations produced by the remaining subjects and plotted in Figure 4 with averages for each subject and image. The PRI was also computed for each of the control segmentations and also included in the same table. The average PRI for each subject was 0.93, which was much higher than the average 0.53 PRI for the random segmentation and the average 0.78 PRI for the ellipse segmentation.

The result of these studies is that the segmentations are remarkably similar across subjects and images, despite the quality of the images, lack of context or semantic understanding, and the specification of just a box. All images, subject segmentations, control segmentations and MATLAB code for computing the above analysis are available online at <http://www.cns.bu.edu/~lgrady/box-study/box-study.html>. Having established the consistency of the segmentations across subjects, we can now proceed to determine how well an algorithm defined using modern techniques can be used to achieve the target segmentation for each image.

## 3. An algorithm for box segmentation

Image segmentation is a very big topic in computer vision which has received a tremendous amount of attention. Since early work on active contours and variational techniques, a prominent approach to image segmentation has been to produce an energy functional, often consisting of multiple terms, to describe the segmentation problem such that lower energy represents a higher quality segmentation. Given this formulation, the segmentation problem may be cast as an optimization problem to find the segmentation with lowest energy. Early work on this topic [16, 22] set the direction for the terms which are most commonly used to describe a segmentation, such as intensity (color, texture) homogeneity within a segmented object, intensity (color, texture) smoothness within a segmented object, short boundary length of the segmented area and low curvature of the boundary of the segmented area. Although these terms have been used to guide reasonable segmentations, the appropriate interplay between terms is unclear and it has been difficult to determine.

General-purpose segmentation from a box has received less attention in the literature than other forms of the segmentation problem. The most prominent algorithm with the box interface is GrabCut [23], which is built on top of the graph cuts algorithm [6] and was later extended to drive the segmentation to “roughly fill the box” [18]. In GrabCut, the authors establish background seeds outside the box, optimize a Gaussian mixture model (GMM) of the separation of color profiles for object/background and use this model to define unary terms for a segmentation.

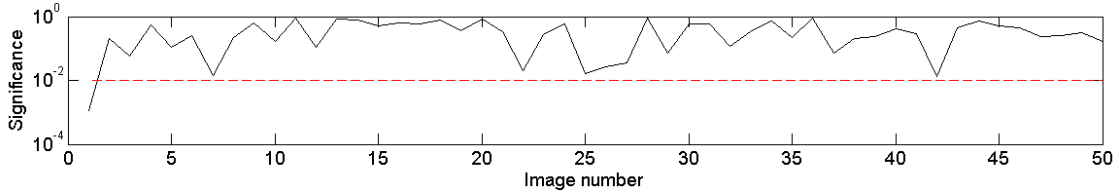


Figure 3. The distribution means of differences of subject segmentations for each image are not significantly different from zero, with a  $p = 0.01$  significance level, as shown on this log plot of the significance levels. The only exception is the first image, in which subjects tended to err on the side of undersegmenting the mean segmentation (although the mean distribution of differences for this image is also not significantly different from zero if a Bonferroni correction is applied). The red dashed line indicates the cutoff for significance.

GrabCut then employs an iterative process of estimating the foreground/background regions using graph cuts and re-estimating the GMM model. One concern about the GrabCut algorithm in the context of our database is that GrabCut assumes a substantial level of separability between the color distribution inside the object and in the background. Although this assumption may be true in many cases, it is by no means always true (as seen in Figure 1).

Our segmentation approach is comprised of two components. The first component of our functional is the ratio of boundary length to volume, i.e.,

$$E(S) = \frac{\partial S}{\text{Volume}(S)}, \quad (8)$$

where  $S$  represents a set of pixels,  $\partial S$  represents the boundary length of the set and  $\text{Volume}(S)$  is assumed to be less than half the volume of all pixels in the image. The ratio in (8) is known as the *isoperimetric ratio* which has been used many times in computer vision, despite the fact that optimization of the isoperimetric ratio is known to be NP-Hard [14]. We formulate our algorithm on a graph,  $G = (V, E)$ , consisting of a node set,  $V$ , associated with the image pixels and an edge set,  $E$ , which we assume here are the edges of a 8-connected lattice. Each edge,  $e_{ij}$ , between nodes  $v_i$  and  $v_j$  is associated with a nonnegative weight  $w_{ij}$ . Consequently, in (8),  $S \subset V$ . In the language of graph theory, one may rewrite (8) in terms of an indicator function,  $x_i$ , which labels node  $v_i \in V$  as belonging to object or background. Since  $x$  indicates membership in  $S$ , we may rewrite (8) as

$$E(x) = \frac{x^T L x}{x^T d}, \quad (9)$$

where  $L$  is the graph Laplacian matrix defined by

$$L_{ij} = \begin{cases} d_i & \text{if } i = j, \\ -w_{ij} & \text{if } \exists e_{ij} \in E, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

such that  $L_{ij}$  is indexed by vertices  $v_i$  and  $v_j$ , and  $d$  is the vector of node degrees,  $d_i = \sum_{e_{ij} \in E} w_{ij}$ .

A method for optimizing (8) was given in [14], in which  $x$  was relaxed to take real values and some nodes were designated as background, i.e., fixed to  $x_i = 0$ . In our case,

these background designations fit well with the box interface, since the box perimeter defines a natural background. As in [14], we may find a solution to optimize (9) by solving the linear system defined by

$$Lx = d, \quad (11)$$

in which the equations corresponding to the fixed nodes are removed (see [14] for more details). The real-valued  $x$  may be converted into a binary partition by choosing a threshold of  $x$  to produce a binary partition that minimizes (9).

The second component of our functional is the probability density function for the object region,  $P$ . Unlike GrabCut, this probability density function is not applied as a unary term, since application as a unary term might create disconnected objects in the image, a target object of higher-order genus and/or fail to distinguish background regions with a similar appearance. Instead, we utilize this probability density to estimate the location of transitions from pixels which have a high probability to be object pixels to pixels which have a low probability to be object pixels.<sup>1</sup> Furthermore, [5, 27] suggested that graphs with undirected edges (and weights) may suffer from inconsistent boundary polarity (i.e., an inconsistent jumping from *interior* boundaries to *exterior* boundaries). Consequently, we adopt a directed graph model in which we penalize less those boundary transitions from pixels with high object likelihood to pixels with low object likelihood than the reverse. Specifically, we define the edge weights as

$$w_{ij} = \begin{cases} \exp -\beta_0 (P(g_i) - P(g_j))^2 & \text{if } P(g_i) > P(g_j) \\ \exp -\beta_1 (P(g_i) - P(g_j))^2 & \text{else,} \end{cases} \quad (12)$$

where  $g_i$  represents the image intensity (grayscale) value at pixel  $v_i$ . In general,  $\beta_0 \gg \beta_1$ , meaning that edges with the correct polarity (from high object probability to low object probability) are penalized much less than the reverse.

<sup>1</sup>We estimate the probability density function for the object rather than the probability density function for the background because, following previous literature [13], we assume that the object is more likely to be homogeneous than the background.

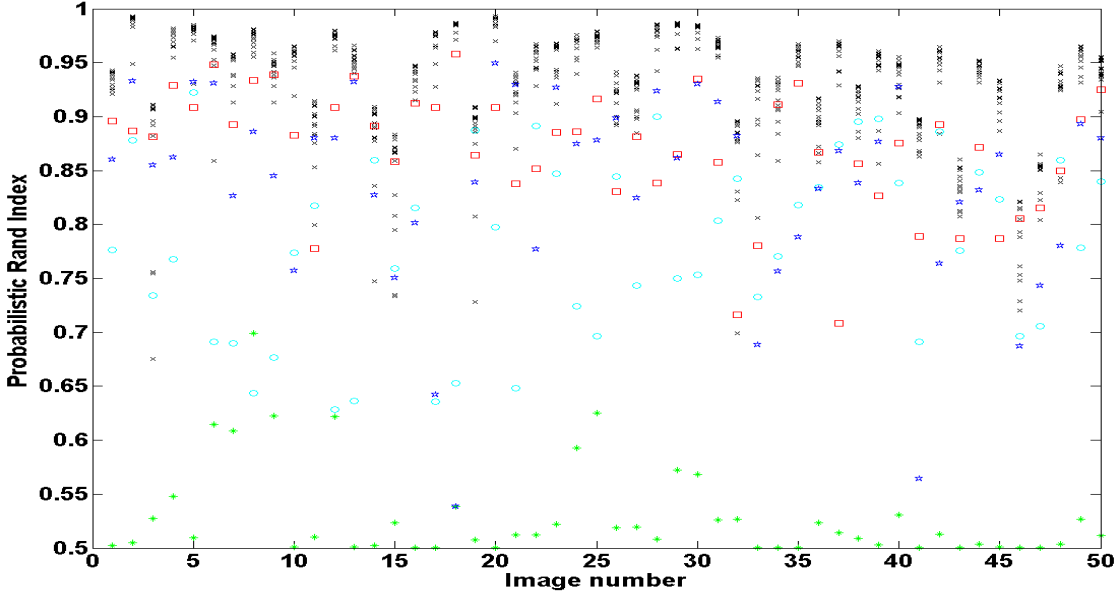


Figure 4. The Probabilistic Rand Index (PRI) for each image. The PRI for each subject (black ‘x’) was measured against the set of other subjects, while the PIBS algorithm (red squares) and GrabCut (blue stars) were measured against the set of all subjects. The control segmentations of an ellipse (cyan circles) and random segmentations (green asterisks) were also measured against the set of all subjects. The average PRI across all images was — Subjects: 0.93, PIBS: 0.87, GrabCut: 0.83, Ellipse: 0.78, Random: 0.53.

We may put together these pieces and write the functional for our box algorithm as

$$E(P, x) = \frac{x^T L_P x}{x^T d}, \quad (13)$$

where  $L_P$  is used to indicate that the weights defining  $L$  are a function of  $P$ . The functional in (13) is optimized by alternating the optimization of  $x$  and the estimation of  $P$  until convergence is achieved (measured by change in the solution between iterations). The solution of a quadratic energy functional with directed weights was addressed in [27] and solved via an iterative minimization, which is the approach we also adopt to minimize (13), given a  $P$ . In alternating iterations,  $P$  is estimated using a standard Parzen window with Gaussian kernel, in which the kernel width is determined from the variance of the intermediate segmentations. To initialize the solution, an  $x$  is calculated by solving (11) using undirected edges and  $P(g_i)$  estimated from the background pixels at the border of the box. To reflect the pieces of our algorithm we term it the Probabilistic Isoperimetric Box Segmentation algorithm (PIBS). Note that PIBS could easily be extended for color, texture or more sophisticated appearance models simply by adjusting the estimation of  $P(g_i)$  for a pixel  $v_i$ .

Figure 4 displays the PRI for the PIBS algorithm on this database, as compared against the 14 subjects, the extended GrabCut algorithm<sup>2</sup> in [18] and the two control segmentations. The average PRI of the PIBS algorithm was 0.87,

<sup>2</sup>The authors thank Victor Lempitsky for providing results on our

which does not achieve the performance of the human subjects, but is substantially better than GrabCut (0.83) and the control segmentations (0.78 and 0.53) for this challenging segmentation task.

## 4. Conclusion

In this paper, we examined the tractability of the problem of creating a general-purpose segmentation algorithm from a box that relies on low-level image features. A major issue in this work was to look for those conditions which appear to be necessary to make the segmentation problem well-posed enough to create an algorithm of measurable quality.

Based on the experiments in Section 2, we conclude that the problem of producing a general-purpose, low-level segmentation from a box is tractable in the sense that the segmentations given by 14 subjects for this difficult task were remarkably consistent. Furthermore, this consistency could not be explained by a random segmentation or a strictly geometric solution (an ellipse). The tractability of the box problem is significant because it provides a simple user interface for interactive segmentation algorithms (which are used in many commercial applications) and because many of the learning-driven object detection algorithms return a box around the target object. Consequently, a general-purpose box segmentation algorithm may be connected with

database from his paper on GrabCut with bounding box priors [18]. GrabCut results were generated with multiple parameter settings and the best scores are reproduced in Figure 4.

any of these detection algorithms to produce an automatic segmentation of a target (learned) object.

A new image segmentation algorithm for the box problem was proposed in this paper which combined the polarity driven intensity models of [27] with the isoperimetric criterion proposed by [14]. On this database, our algorithm compared favorably with the extended GrabCut algorithm in [18], likely because of the image monochromaticity and the significant overlap between the appearance of object and background. The PIBS segmentation algorithm performed significantly better than the control segmentations, but not quite at the performance level of the human subjects. The database images and subject segmentations are available online at <http://www.cns.bu.edu/~lgrady/box-study/box-study.html>, which will allow other research groups to leverage this research to perfect a box-driven segmentation algorithm and therefore solve this increasingly important problem.

## References

- [1] B. Appleton and H. Talbot. Globally optimal surfaces by continuous maximal flows. *PAMI*, 28(1):106–118, Jan. 2006. **1**
- [2] P. Arbelaez. Boundary extraction in natural images using ultrametric contour maps. In *Proc. of POCV*, 2006. **1**
- [3] T. Athanasiadis, P. Mylonas, Y. Avrithis, and S. Kollias. Semantic image segmentation and object labeling. *IEEE Trans. on Circuits and Systems*, 17(3):298–312, 2007. **3**
- [4] X. Bai and G. Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. In *Proc. of ICCV*, 2007. **1**
- [5] Y. Boykov and G. Funka-Lea. Graph cuts and efficient n-d image segmentation. *IJCV*, 70(2):109–131, 2006. **6**
- [6] Y. Boykov and M.-P. Jolly. *Interactive graph cuts* for optimal boundary & region segmentation of objects in N-D images. In *Proc. of ICCV 2001*, pages 105–112, 2001. **1, 2, 5**
- [7] J. S. Cardoso and L. Corte-Real. Toward a generic evaluation of image segmentation. *IEEE TIP*, 14:1773–1782, 2005. **1**
- [8] A. X. Falcão, J. K. Udupa, S. Samarasekera, S. Sharma, B. H. Elliot, and R. de A. Lotufo. User-steered image segmentation paradigms: Live wire and live lane. *Graphical Models and Image Processing*, 60(4):233–260, 1998. **4**
- [9] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004. **1**
- [10] E. Fowlkes and C. Mallows. A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78(383):553–569, Sep. 1983. **1**
- [11] B. V. Ginneken, T. Heimann, and M. Styner. 3D segmentation in the clinic: a grand challenge. In *Proc. of MICCAI Workshop*, pages 7–15, 2007. **1**
- [12] L. Grady. Random walks for image segmentation. *IEEE PAMI*, 28(11):1768–1783, 2006. **1**
- [13] L. Grady and M.-P. Jolly. Weights and topology: A study of the effects of graph construction on 3D image segmentation. In D. M. et al., editor, *Proc. of MICCAI 2008*, volume Part I of LNCS, pages 153–161. Springer-Verlag, 2008. **6**
- [14] L. Grady and E. L. Schwartz. Isoperimetric graph partitioning for image segmentation. *IEEE PAMI*, 28(3):469–475, March 2006. **1, 6, 8**
- [15] M.-P. Jolly, C. Alvino, B. Odry, X. Deng, J. Zheng, M. Harder, and J. Guehring. Automatic femur segmentation and condyle line detection in 3D MR scans for alignment of high resolution MR. In *ISBI*, pages 940–943, 2010. **2**
- [16] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1988. **3, 5**
- [17] H. I. Koo and N. I. Cho. Rectification of figures and photos in document images using bounding box interface. In *Proc. of CVPR*, pages 3121–3128, 2010. **2**
- [18] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In *Proc. of ICCV*, pages 277–284, 2009. **2, 4, 5, 7, 8**
- [19] Y. Li, J. Sun, C. Tang, and H. Shum. Lazy snapping. In *Proc. of SIGGRAPH*, volume 23, pages 303–308, 2004. **2**
- [20] 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 *Proc. of ICCV*, 2001. **1**
- [21] E. Mortensen and W. Barrett. Interactive segmentation with intelligent scissors. *GMIP*, 60(5):349–384, 1998. **4**
- [22] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure and Appl. Math.*, 42:577–685, 1989. **3, 5**
- [23] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *Proc. of SIGGRAPH*, volume 23, pages 309–314, 2004. **2, 5**
- [24] J. Santner, T. Pock, and H. Bischof. Interactive multi-label segmentation. In *Proc. of ACCV*, Nov. 2010. **2**
- [25] J. Sethian. *Level set methods and fast marching methods*. Cambridge University Press Cambridge, 1999. **2**
- [26] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE PAMI*, 22(8):888–905, Aug. 2000. **1**
- [27] D. Singaraju, L. Grady, and R. Vidal. Interactive image segmentation of quadratic energies on directed graphs. In *Proc. of CVPR 2008*. IEEE, June 2008. **6, 7, 8**
- [28] A. K. Sinop and L. Grady. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In *Proc. of ICCV 2007*, Oct. 2007. **2**
- [29] A. Torralba, K. Murphy, and W. Freeman. Sharing visual features for multiclass and multiview object detection. *IEEE PAMI*, pages 854–869, 2007. **2**
- [30] M. Unger, T. Pock, W. Trobin, D. Cremers, and H. Bischof. TVSeg - Interactive total variation based image segmentation. In *Proc. of BMVC*, 2008. **1**
- [31] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE PAMI*, 29(6):929–944, June 2007. **1, 4, 5**
- [32] L. Vese and T. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *IJCV*, 50(3):271–293, 2002. **3**
- [33] P. Viola and M. Jones. Robust real-time object detection. *IJCV*, 57(2):137–154, 2002. **2**
- [34] P. Wighton, M. Sadeghi, T. K. Lee, and M. S. Atkins. A fully automatic random walker segmentation for skin lesions in a supervised setting. In *MICCAI*, pages 1108–1115, 2009. **2**