

Comparative Evaluation of Interactive Segmentation Approaches

Mario Amrehn¹, Jens Glasbrenner¹, Stefan Steidl¹, Andreas Maier^{1,2}

¹Pattern Recognition Lab, FAU Erlangen-Nürnberg

²Erlangen Graduate School in Advanced Optical Technologies (SAOT)

mario.amrehn@fau.de

Abstract. Image segmentation is a key technique in image processing with the goal to extract important objects from the image. This evaluation study focuses on the segmentation quality of three different interactive segmentation techniques, namely Region Growing (RG), Watershed (WS) and the cellular automaton based GrowCut (GC) algorithm. Three different evaluation measures are computed to compare the segmentation quality of each algorithm: Rand Index (RI), Mutual Information (MI), and the Dice Coefficient (D). For the images in the publicly available ground truth data base utilized for the evaluation, the GrowCut method has a slight advantage over the other two. The presented results provide insight into the performance and the characteristics with respect to the image quality of each tested algorithm.

1 Introduction

Image segmentation is the process of extracting relevant objects from the image on the basis of local image features such as texture features or pixel intensities. In the case of medical imaging such objects may be lesions, bones, vessels, or other coherent structures. Since there are numerous types of circumstances which can deteriorate the image quality significantly, the segmentation process is a challenging task, depending on the type of image data, the image quality, and the characteristics of the object to segment. In general, image segmentation can be classified into fully automatic segmentation and interactive segmentation. The former requires a large amount of prior knowledge about the characteristics of the target object as well as the image background, which is often not available. In interactive segmentation frameworks parts of the object and the background are roughly indicated by human operators by means of corresponding markers, so-called seeds. By providing some additional seeds after the initial segmentation result is obtained the user is able to adjust the segmentation quality until satisfaction. There are typical problems when dealing with medical images, such as low contrast, image noise and artifacts caused by the imaging device or reconstruction algorithm. In cases where the objects to segment have a high diversity, e. g. no typical appearance and no strong intensity boundaries, a fully automatic segmentation approach would most probably fail to produce high quality segmentations. To overcome these problems a semi-automatic approach guided by

human user interaction is applied. Since humans usually outperform computers in object recognition [1] an interactive approach can provide useful information about the size and the location of the target object. The segmentation quality of three different user-guided image segmentation approaches on high quality 2D images is explored, namely Watershed [2], Region Growing [3] and GrowCut [4].

2 Materials and methods

2.1 GrowCut

The GrowCut algorithm is based on cellular automaton theory. A cellular automaton is a quad-tuple (Z^n, S, N, δ) , where $Z^n \ni p$ is the pixel or cell space, n is the image's dimensionality and $S \ni S_p^t = (l_p^t, \Theta_p^t, c_p)$ is a non-empty state set. l_p^t is the label (foreground (FG), background (BG) or undefined (UN)) of cell p at time t , Θ_p^t its strength and c_p is the cell's feature vector. Usually the neighborhood system N in 2D is either the von Neumann 4-pixel neighborhood or the Moore 8-pixel neighborhood to define the graph's edges. Each iteration $\delta_t(S_p^t) = \delta_t(l_p^t, \Theta_p^t, c_p) = (l_p^{t+1}, \Theta_p^{t+1}, c_p) = S_p^{t+1}$ is performed with δ as a local state transition rule. The user marks some pixels of the object to segment as FG and at least one BG pixel. Starting from these labeled pixels, the GrowCut algorithm propagates the labels based on local intensity features to compute an optimal segmentation. Henceforward, at each discrete time step, each cell tries to attack its neighbors and thus tries to occupy the whole cell space based on its current attacking strength. Since the attacking strength is the cell strength Θ_p^t times g as defined in (1), the process is guaranteed to converge with bounded cell strengths monotonously decreasing

$$[0, 1] \ni g(c_p, c_q) = 1 - \frac{\|c_p - c_q\|_2}{\max_{i,j \in Z^n} (\|c_i - c_j\|_2)} \quad (1)$$

2.2 Watershed

As a pre-processing step prior to the Watershed segmentation a gradient image is computed as depicted in Fig. 2.5. This image, holding information about the edges, is considered an altitude surface in which high intensity pixel values correspond to ridge points while low intensity pixel values correspond to valley points. If a drop of water drips on a high-altitude point, it flows down the surface into a local minimum, a catchment basin. The final segmentation is composed of all watersheds, whereby all pixels that belong to the same watershed are added to the same segment in the image. The classical Watershed algorithm is an unsupervised segmentation procedure. In order to provide user control, the topography of the precomputed edge image is modified with a marker image containing the manually drawn FG and BG seeds, such that unwanted boundaries are removed [5].

2.3 Region growing

In the Region Growing algorithm the initially drawn FG seeds expand over the whole image until a user-defined stop canon is fulfilled. Iteratively all neighbors of each pixel are explored and assigned a label according to a predefined threshold. In the presented case the considered unlabeled pixel will not be added to the region if the difference between the pixel's intensity value and the mean value of all pixel's within that region is above a certain threshold. This method is included as a baseline in this segmentation quality evaluation.

2.4 User interface

In order to test the algorithms efficiently and to provide an intuitive user experience an interactive user interface was implemented. Fig. 3 shows a screenshot of the tool. The work flow is as follows: first the user draws the Region of Interest, which is indicated by a blue rectangle. This ROI provides some prior knowledge by roughly indicating the position and size of the object to segment. The user selects one of the previously introduced algorithms. Finally, the user is able to draw the FG and BG seeds respectively by dragging and clicking the mouse over the corresponding image regions. After the first segmentation result is computed the result is presented to the user by indicating the object boundary superimposed on the original image as shown in Fig. 2.5. In order to obtain a satisfying segmentation quality it is possible to reedit the result by adding additional foreground or background seeds. The segmentation algorithm is run again based on the previously computed segmentation result. This step can be repeated until the user is satisfied with the segmentation quality.

2.5 Metrics

Obtaining ground truth data is a tedious and time consuming process and is subject to operator variability. In order to perform a quantitative evaluation of the segmentation outcome the Microsoft Research ground truth database [6] containing 30 quality images and ground truth segmentations is used. Each computed mask is evaluated against the manually segmented ground truth using the following well-known segmentation quality metrics: the Rand Index [7], the Mutual Information [8], and the Dice Coefficient [9]. The segmentations are obtained with the presented segmentation tool. In order to guarantee a fair comparison, the same seeds are used for all algorithms for each respective image. Note that the threshold for the Region Growing algorithm was optimized for each image in order to obtain the best individual RG segmentation result as a baseline.

Rand index The Rand Index is a similarity measure between two clusterings. It represents the percentage of pairs of pixels where the computed segmentation and the ground truth are the same. Let s be the number of pairs of pixels that

are in the same object in segmentation A and in segmentation B and d the number of pairs of pixels which are in different objects in the two segmentations. Then the Rand Index is computed by $ri(s, d) = (s + d) / \binom{n}{2}$ where n is the number of pixels in the image. The adjusted Rand Index is adjusted-for-chance to guarantee that the measure is bounded between 0.0 and 1.0. A value of 0.0 indicates a completely different segmentation result and a value of 1.0 indicates a perfect match between the computed segmentation and the ground truth.

Mutual information The MI is a probabilistic measure for the mutual dependence between the ground truth and the computed segmentation. A normalized MI of 1.0 between two clusters indicates a perfect agreement of the two clusters, since the segmentation and the ground truth share the maximum amount of information. The MI of two sets of clusters $A \ni A_i$ and $B \ni B_j$ is computed by (2), where $P(A_i, B_j)$ denotes the joined probability that a point belongs to both clusters A_i and B_j . $P(c)$ is the probability of a random label occurring in cluster c

$$[0, 1] \ni mi(A, B) = \sum_{A_i \in A} \sum_{B_j \in B} P(A_i, B_j) \log \left(\frac{P(A_i, B_j)}{P(A_i)P(B_j)} \right) \quad (2)$$

Dice coefficient The Dice metric [9] calculates the spatial overlap coefficient of two segmentations A and B. Calculated by $d(A, B) = 2 \cdot (|A \cap B|) / (|A| + |B|)$, a perfect match of the computed segmentation with the ground truth is scored 1.0, while a value of 0.0 indicates a completely different labeling.

3 Results

Tab. 1 shows the Adjusted Rand Index, Mutual Information and Dice Coefficient values averaged over all 30 test images with respect to the ground truth data for each algorithm. Fig. 3 shows a selection of some segmentation examples for each algorithm as well as the corresponding ground truth.

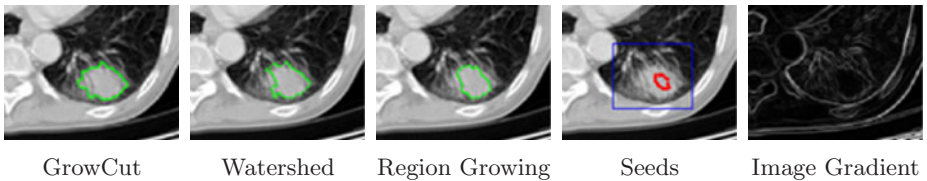


Fig. 1. Segmentations using the same set of FG and BG seeds for each algorithm. The inner red labels represent FG seeds, the ROI is outlined by a blue rectangle. Source image obtained from [10].

Table 1. Average segmentation results using ARI, MI, D metrics for each algorithm.

Segmentation	Adjusted Rand Index	Mutual Information	Dice Coefficient
GrowCut	93.72%	86.88%	96.14%
Watershed	92.73%	85.01%	95.49%
Region Growing	42.59%	32.48%	57.70%

4 Discussion

The results presented in Tab. 1 as well as in Fig. 3 show that each algorithm has its advantages and drawbacks depending on the quality of the underlying image data. However, the GrowCut segmentation algorithm slightly outperforms the other tested segmentation methods on this particular data set, while the Region Growing algorithm performed clearly worse.

The Region Growing segmentation approach has the advantage that it is easy to implement, has a low computational complexity and is capable of creating satisfying segmentation results for images with clear object boundaries as one can see in the first two rows on the right side of Fig. 3. However, a major drawback of this approach is that the threshold needs to be adjusted carefully to the underlying image properties in order to obtain a satisfying segmentation. In addition, the growing FG region tends to flood neighboring segments in the presence of noise or weak intensity boundaries as depicted in the third row.

The GrowCut algorithm is able to achieve quality segmentations, but in some images the resulting segmentation boundary can be ragged. One can see this effect in the third row on the right side of Fig. 3. This may be acceptable or even

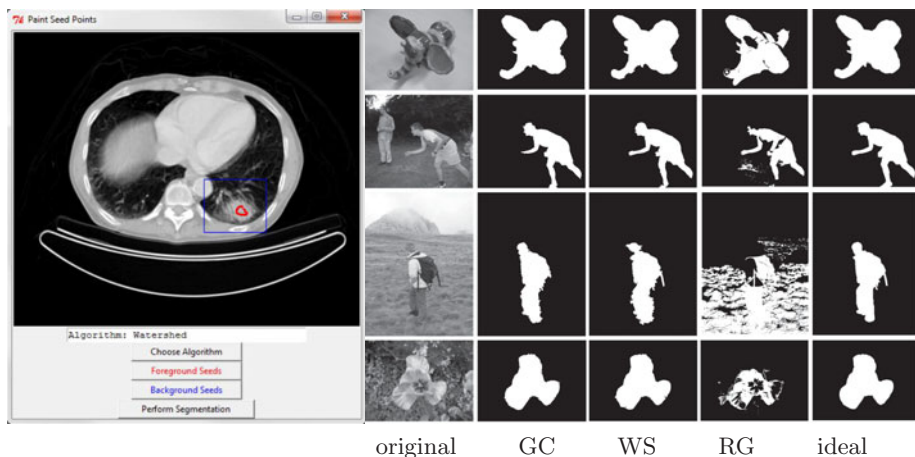


Fig. 2. Left: User Interface of the segmentation framework. The user defined region of interest is indicated by a blue rectangle. The red dots represent the FG seeds which can be drawn using a brush stroke. Right: Segmentation results for images from the ground truth database [6]. Columns from left to right: Original image, GrowCut segmentation, Watershed segmentation, Region Growing segmentation, ground truth.

desired if the segmentation boundary should be very detailed. However, this can also be an undesired effect when the objects to segment typically do not have ragged surfaces, e. g. human structures like bones or organs. A morphological post-processing of the binary segmentation mask might be performed to generate smoother object boundaries.

Since the Watershed algorithm operates on a pre-processed gradient image it is sensitive to false or unclear edges in the image. If too many edges are detected the watershed algorithm tends to produce an over-segmentation which can be seen in the third row on the right side of Fig. 3. Here, the texture of the background in the original image produces an ambiguous gradient image which results in a jagged segmentation boundary in these regions.

In general, all tested algorithms are able to produce quality segmentations depending on the object to segment and the underlying image data. Nevertheless, a user-guided image segmentation is always to some extent subject to the operator variability and depends on the operator's effort to produce a satisfying segmentation outcome. This can intuitively be explained by the fact that less seeds mean less information which is provided to the algorithm and vice versa. Consequently, any interactive segmentation framework is always a trade-off between its efficiency with respect to the user's effort and the segmentation quality.

References

1. Spina TV, Falcão AX, Miranda PAV. User-steered image segmentation using live markers. *Comput Anal Images Patterns*. 2011; p. 211–8.
2. Beucher S, et al. The watershed transformation applied to image segmentation. *Scanning Microsc Suppl*. 1992; p. 299.
3. Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell*. 1994;16(6):641–7.
4. Vezhnevets V, Konouchine V. GrowCut: interactive multi-label ND image segmentation by cellular automata. *Proc Graphicon*. 2005; p. 150–6.
5. Beare R, Lehmann G. The watershed transform in ITK-discussion and new developments. *Insight J*. 2006.
6. Rother C, Kolmogorov V, Blake A. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans Graph*. 2004;23(3):309–14.
7. Rand WM. Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc*. 1971;66(336):846–50.
8. Russakoff DB, Tomasi C, Rohlfing T, et al. Image similarity using mutual information of regions. *Proc ECCV*. 2004; p. 596–607.
9. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26(3):297–302.
10. Rodt T, von Falck C, Halter R, et al. Mikro-CT basierte region growing segmentation zur verlaufsbeurteilung des multifokalen adenocarcinoms der lunge in SPC-raf transgenen mäusen. *Rofo*. 2010;182:WLPO5.