

How Interaction Methods Affect Image Segmentation: User Experience in the Task

Ramya Hebbalaguppe, Kevin McGuinness, Jogile Kuklyte, Graham Healy
Noel O'Connor and Alan Smeaton
Center for Digital Video Processing, CLARITY
Dublin City University

*ramya.hebbalaguppe2@mail.dcu.ie

Abstract

Interactive image segmentation is extensively used in photo editing when the aim is to separate a foreground object from its background so that it is available for various applications. The goal of the interaction is to get an accurate segmentation of the object with the minimal amount of human effort. To improve the usability and user experience using interactive image segmentation we present three interaction methods and study the effect of each using both objective and subjective metrics, such as, accuracy, amount of effort needed, cognitive load and preference of interaction method as voted by users. The novelty of this paper is twofold. First, the evaluation of interaction methods is carried out with objective metrics such as object and boundary accuracies in tandem with subjective metrics to cross check if they support each other. Second, we analyze Electroencephalography (EEG) data obtained from subjects performing the segmentation as an indicator of brain activity. The experimental results potentially give valuable cues for the development of easy-to-use yet efficient interaction methods for image segmentation.

1. Introduction

Image segmentation is an important yet still unsolved problem in computer vision. There are three main approaches to the task: automatic, manual, and interactive. Manual segmentation is time-consuming and labour intensive, and issues like scene complexity, ambiguous objects of interest, and low contrast often cause fully automatic methods to fail. Intervention of a human operator in image segmentation is often termed *semi-supervised* or interactive image segmentation. A wide range of applications including image editing, biomedical image analysis, and digital image composition motivate the development of efficient interaction methods for image segmentation. Published research in

interactive image segmentation generally places greater emphasis on algorithms and their optimization than on interaction methods; representative work that deal primarily with algorithms for interactive image segmentation can be found in [1, 4, 7, 14]. User experience during interactive image segmentation is undoubtedly important, but comparatively little attention has been devoted to the development of interaction methods that take into account usability and user experience [13]. JetStream [12] provides a robust approach to extracting regions of interest but demands a large amount of attention from the user. Although [5] introduce a lazy snapping tool that focuses on improving user experience, and [7] emphasize the desirability of a small amount of user input for mobile applications, they only evaluate their approaches using measures of accuracy. A different approach to interactive segmentation is introduced in [9] where eye gaze is used as a mode of interaction to place seeds for image segmentation. Accurate placement of seeds, however, can be strenuous on the eyes when a complex natural image is presented.

We focus our research on a study of the effects of interaction methods on image segmentation for photo-manipulation purposes, while leveraging existing research in comparing image segmentation algorithms for accuracy, efficiency, and repeatability. Our study considers the importance and trade-offs between stressful interaction methods and segmentation accuracy. To evaluate user experience while segmenting images, in addition to the usual accuracy measures, we evaluate the amount of user interactions, the average time taken, and number of strokes per segmentation. We also analyze EEG signals in an attempt to measure brain activity and stress levels, and describe a survey aimed at achieving a better understanding of the user experience.

We use the interactive graph cuts (IGC) image segmentation algorithm as it was previously found to be preferred by most users and to perform on par with the best algorithms in terms of accuracy [11]. Our experiments are performed

Table 1. Interaction Methods used for experiments

| Code | Method | Description |
|------|----------------------------|--|
| M1 | Scribble | Subject marks regions of interest with mouse strokes – foreground with the left mouse button and background with the right |
| M2 | Bounding box with scribble | An initial bounding box is drawn around the region of interest to improve the speed of segmentation. The subsequent steps are as in M1 |
| M3 | Outline with scribble | Free form drawing around the boundary of object followed by scribble for refinement of object |

on a set of randomized images taken from the Berkeley segmentation dataset [10]. To facilitate experiments, an interactive segmentation tool with various interaction methods was developed. A demonstration of the interaction methods implemented for this study is available online¹. The interactive image segmentation software is also available for download².

2. Interaction Methods

Interactive image segmentation requires both *object recognition* to indicate object location in the image and *object delineation* to define its precise spatial extent. As humans usually outperform computers in recognition [15], it is desirable to have interaction methods that combine automatic techniques with human recognition abilities having minimal effort from the user. In this study we analyze the effects of different *interaction* methods while keeping the *object delineation* approach constant. Table 1 lists the interaction methods developed for our study. The most intuitive is *scribble* where the initial step is to roughly indicate background and foreground regions using mouse strokes. When at least one stroke for each segment is completed, the program makes a prediction based on user input and displays it visually on screen. The user analyzes the visual information and refines the object accordingly, refining until the user is satisfied with the segmentation. The second method is based on the assumption that much background noise can be eliminated by localization of the region of interest to improve the speed at which the segmentation is performed. After the user draws a bounding box around the boundary of the object, the *scribble* method is used to refine the segmentation. In some extreme cases, such as images where object regions are highly textured and do not exhibit region homogeneity or images where boundaries are weak or intermittently absent or they exist among strong non-object boundaries, it is difficult for an algorithm to predict the correct object boundaries. Therefore we implement another method, where a user can draw the contour of the object. The *scribble* technique is again used to further refine bound-

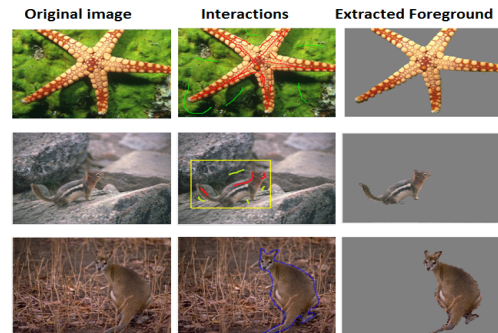


Figure 1. First row: scribble; second row: bounding box + scribble; third row: outline + scribble

ary accuracy. Figure 1 depicts how foreground is extracted using the three interaction methods discussed before.

In this paper we discuss and compare the efficiency of the aforementioned user interaction methods and evaluate their performance in terms of: (a) object accuracy and boundary accuracy; (b) time taken to segment an image using each interaction method; (c) amount of user input required; (d) user opinion about the experience through a survey; and (e) electroencephalography (EEG) signals captured during experiments.

3. Experiments

The evaluation of interaction methods was carried out as a series of user experiments in which subjects were tasked with extracting an object from an image with a time constraint of two minutes. Twenty postgraduate students and research staff from Electronic and Computing Department of a university were selected as participants, comprising 12 male and 8 female, and the ages spanned from 23 to 38 with average 25 years. Their proficiency level was intermediate. Segmentation for each image was constrained to 120 seconds to avoid the users spending too much time refining a segmentation. This can be justified since the primary purpose of interactive segmentation is to provide an accurate segmentation faster than it would take to produce it fully manually. In total, 30 images in a randomized order taken

¹<http://www.youtube.com/watch?v=eMCG3jUOI44>

²<https://dl.dropbox.com/u/6200076/InterImgSegSoftware.zip>

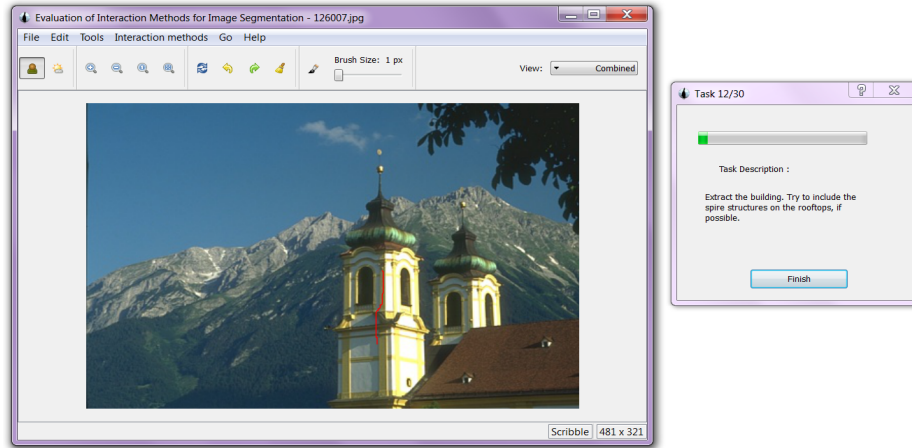


Figure 2. Screenshot while evaluation of interactive segmentation tool in experiment mode



Figure 3. Views supported by interactive image segmentation software. First row: left, *markup view* and right: *mask*. Second row: left, we have *Outline overlaid* mode and on the right, we have *Foreground only* mode

from the Berkeley Segmentation Dataset [10] were used – 10 images per interaction method and the duration of experiments spanned one hour in total. As participants refined their respective segmentations, the updated segmentation masks were stored along with the timestamps forming a progressive collection of segmentations over time and a corresponding time-stamp in a log file as new refinements are added. The screenshot of the software in experiment mode can be seen in Figure 2. The application also supports different views to enable the users to accurately extract the object, some of the views supported are shown in Figure 3.

Each recorded mask was evaluated against a manually segmented ground-truth, thus allowing us to gauge segmentation accuracy over time. We also estimated user input in terms of average time taken to segment each image and in

terms of number of strokes required to complete the segmentation from the time stamps. In addition, we captured EEG data during all the experiments to gain insight into the stress level of the user during the process.

There were two sets of experiments conducted: one without EEG and one with an EEG setup. 14 users participated in the experiments without EEG and 6 subjects participated in those with EEG owing to time overhead of setting up low-cost EEG equipment.

4. Results

A set of subjective and objective metrics were obtained that measure both usability and user experience. These indicators measure human performance and user satisfaction in achieving both pragmatic and hedonic goals.

The subjective parameters are (a) user preference for the interaction methods and (b) user rating of each interaction method used for segmenting an object. The objective metrics include (a) the time taken to segment an image using each interaction method; (b) object accuracy and boundary accuracy; (c) EEG measurements; and (d) the amount of user input required. For the subjective metrics, standard tests for significance are performed on data listed in tables 2, 4 and 5. We chose the non parametric Wilcoxon test to evaluate the difference between paired samples to determine if statistical assesment of mean of ordinal values are harmonic with the individual tests [8]. The results obtained conform to the requirements of statistical significance of data obtained by subjective metrics.

4.1. User Preference Survey

The subjects were asked to vote for a single interaction method among the three methods discussed in Table 1. Table 2 shows the votes received by each. The table shows that

based on performance of segmentation and ease with which an interaction method could be used for segmentation, the bounding box+scribble (M2) was the most-preferred interaction, followed by the scribble (M1). Since user preference constrained the users to vote for a single interaction method, users ratings were also collected on each interaction method using a five-point Likert scale as described in next section.

4.2. User Ratings

User ratings were measured using a Likert scale (Table 3). The Likert scale is commonly used in surveys as it allows the subjects to quantify opinion based items [6]. The subjects specify their level of agreement or disagreement on a symmetric agree-disagree scale for a series of statements. In our experiment, the Likert scale was used measure opinion on two questions/statements: (a) I found the interaction method to be easy to use; (b) I found the interaction method to be useful. Table 4 shows the mean Likert scale ratings over 20 subjects for the three interaction methods. We note that scribble was easiest for subjects to use when compared to *bounding box+scribble* and *outline methods*. Overall, we can conclude *bounding box+scribble* is rated well both on usefulness and ease parameters while the outline got lower ratings for both usefulness and ease Likert items. Through the questionnaire and post-experiment informal discussions with individual participants, we found that the drawing a rectangle in case of the *bounding box+scribble* method seemed to provide a quick initial segmentation 'kickstart', skipping a tedious scribbling at the early stage of the process. This skipping of the first few strokes of scribbling seems to be the reason for the participants' declared preference.

Table 2. User Preference: votes for each interaction method

| Method | Votes |
|-----------------------|-------|
| Scribble | 6 |
| Bounding Box+Scribble | 10 |
| Outline+Scribble | 4 |

Table 3. Likert Scale for interaction methods (Range:1-5)

| Rating | description |
|--------|-------------------|
| 1 | strongly disagree |
| 2 | disagree |
| 3 | neutral |
| 4 | agree |
| 5 | strongly agree |

4.3. Required Time for User Input

The time stamps obtained from the log file were analysed to measure average time taken to segment an image and also average number of strokes per each interaction method. Table 5 shows the mean time taken to segment, averaged over all images. It is clear that the *bounding box + scribble* method takes the least amount of time, followed by the outline method. Drawing an outline, however, proved strenuous for the user.

4.4. Average number of strokes

We observed a correlation between user preference and the amount of input the user provides. Minimal interaction was preferred by all users. We observed that, on average, users made 26 strokes to extract an object using the *scribble* method. Using the *bounding box+scribble* method, the number of strokes was reduced by an average of 5 over *scribble*. Using the *outline+scribble* method, the number of strokes was reduced by an average of 7 over *scribble*. The number of strokes were minimum for the *Outline+scribble* method: the first stroke drawn is often quite accurate and less effort is needed for refinements.

4.5. Object and Boundary Accuracies

In addition to the final segmentation result, the segmentation masks after each user interaction were assessed. This allows us to analyze the progress of the object extraction using different interaction methods. Masks were evaluated against the ground-truth dataset from [11]. The Jaccard index was used for measuring object accuracy, and a fuzzy metric [11] was used to measure boundary accuracy as they both show reasonable correlation between measured and perceived accuracy. Figure 4 shows average object and boundary accuracies for all users.

From Figure 4, It is evident that initial boundary and object accuracy is higher for the *bounding box+scribble* interaction method and the *outline* method. The *outline* method

Table 4. Mean Likert scale ratings

| Method | Usefulness | Ease |
|-----------------------|------------|------|
| Scribble | 3.45 | 4.10 |
| Bounding Box+Scribble | 3.81 | 3.90 |
| Outline+Scribble | 2.45 | 2.18 |

Table 5. Average segmentation time

| Method | Time(secs.) |
|-----------------------|-------------|
| Scribble | 59 |
| Bounding Box+Scribble | 39 |
| Outline+Scribble | 50 |

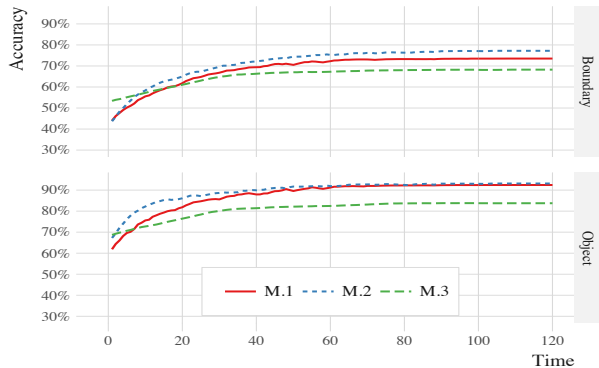


Figure 4. Mean boundary and object accuracy over all users as a function of time

achieves lower final accuracy than the *scribble* or *bounding box+scribble* methods. This implies that the choice of interaction method can effect accuracy even when the underlying image segmentation algorithm is the same. For both boundary and object accuracy, the final value achieved for *bounding box+scribble* was approximately 10% higher than for the *outline+scribble* interaction method. Figure 4 also shows that the *bounding box+scribble* method consistently outperforms the *scribble* interaction method.

4.6. EEG Analysis

EEG (Electroencephalography) involves the detection of electrical potentials generated by the brain on the scalp and is detected by affixing electrodes with a conductive gel to the head. These signals typically display a number of oscillatory components referred to by their frequency bands: delta rhythm (1-3 Hz), theta rhythm (5-7 Hz), alpha rhythm (8-12 Hz), beta rhythm (13-30 Hz), gamma rhythm (above 30 Hz), and mu rhythm (8-13 Hz). Perturbations in these signals are correlated with cognitive processes including those involved with attention, emotion and decision-making. There is growing interest in the potential of these signals to be used for non-medical applications [2] for instance in applications involving performance monitoring in users. A KT88-1016 EEG system was used with electrodes placed at scalp locations Fz, Cz, Pz, Oz, C3, C4, P3, P4 corresponding to the 10-20 electrode placement system. These signals were digitized at 100Hz, and subsequently band-passed from 0.1Hz to 20Hz. The left earlobe was used as the reference, with the chin used as ground.

Feature vectors were extracted from 1.5 second epochs using an FFT capturing between 0-20Hz of size 20. Amalgamating the extracted features for each of the 8 EEG channels resulted in a feature vector comprised of 160 values. To examine and derive a set of measures of the detectability of the presence of EEG signals that differentiate between using the segmentation methods, we used a support vector machine (SVM) with a linear kernel [3]. Using a re-

Table 6. AUC results from classifiers

| Subject | M1 v M2 | M2 v M3 | M1 v M3 |
|---------|---------|---------|---------|
| S1 | 0.77 | 0.63 | 0.67 |
| S2 | 0.59 | 0.62 | 0.77 |
| S3 | 0.68 | 0.81 | 0.71 |
| S4 | 0.61 | 0.65 | 0.69 |
| S5 | 0.54 | 0.65 | 0.74 |
| S6 | 0.69 | 0.67 | 0.72 |
| Average | 0.65 | 0.67 | 0.72 |

peated random sub-sampling validation (20 times) we derive a measure of how well the EEG signals between segmentation methods can be differentiated.

The classification performance between different methods was measured using the AUC (Area Under Curve) of the ROC (Receiver Operating Characteristic) curve where an AUC of 0.5 is considered chance.

4.6.1 EEG Results

Table 6 shows that different patterns of EEG activity are present when using different image segmentation methods, indicating that the EEG signal is measurably influenced by the segmentation task. There were differences in boundary and object accuracy differences among the interaction methods. Comparing average AUCs across users for each method suggests the signals present while using the outline method differ more so than the other two methods do between each other, with the last comparison (M1 v. M3) showing the greatest difference. A reason for this, we suggest that the cognitive load involved in drawing an accurate outline eliciting more distinct activity as compared with the scribble and/or bounding box methods. Significant classification accuracies were achieved in nearly all cases in Table 6. These results signify the presence of signal changes occurring as a result of the user’s cognitive attentional strategy and differences in allocation of attentional resources while using each of the methods. This is not surprising given that users stated preferences of particular methods reflecting factors like their perceived ease/difficulty of use. These subjective differences on the neural level for using different segmentation methods support the presence of the preference differences we see in the user satisfaction feedback survey across users. We find that while the segmentation methods may seem ostensibly similar in their performance, other factors warrant consideration like preference of a method due to its ease of use by that user. By analysing EEG signals with these measures we can assert subjective differences in cognitive demands do exist across methods and are responsible in shaping these preferences.

5. Discussion

We analyzed the metrics that demonstrate improvement of user experience in segmenting objects using various interaction methods. The objective metrics used were (a) average time to segment an image using a particular interaction method; (b) object and boundary accuracies using different interaction methods; (c) Average number of strokes used for refinement; and (d) EEG data to support these metrics. The subject metrics were (a) user preference survey and (b) user ratings of every interaction method on Likert scale.

Analysis of the results showed that the interaction method affects segmentation accuracy. Segmentation for each image was constrained to 120 seconds and object and boundary accuracies were found to be higher using *bounding box+scribble* interaction method. Of course, segmentation accuracy also depends on the segmentation algorithm applied, but in our experiments the segmentation algorithm was constant over all interaction methods. We also noticed that minimizing time taken to segment the image would relate to a better user experience. For example, the average time taken for the user to obtain a satisfactory segmentation when using the *bounding box with scribble* method was observed to be 20 seconds less than the scribble method, and most users preferred this approach. Although the *outline with scribble* method took 9 seconds less on average than the *scribble* interaction method, some subjects reported that the *outline+scribble* method is time consuming and could potentially cause shoulder/arm cramps while drawing along the contour, particularly when a complex object was presented for extraction. Adding a feature where the user can localize segmentation by drawing a bounding box around the object was generally preferred by users. This shows that it is useful to work on features added to a simple interaction approach.

6. Conclusion

This paper investigated the various factors at play in an interactive segmentation process from a user perspective. The experiments performed provide insight into a number of different interaction strategies in terms of their accuracy but also associated user preference and cognitive load. In the future we would like to investigate and develop more interaction methods to extend the scribble approach, including looking at simplifying the drawing of a bounding box as this method gave higher subjective and objective ratings. We would also like to explore if a supposedly lower performing delineation method gives better results if it were to be implemented with *bounding box+scribble*. In conclusion, we found that it is important to offer different intuitive interaction methods, but to let the user choose the most appropriate approach.

References

- [1] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive gmmrf model. In *ECCV*, volume 3021 of *Lecture Notes in Computer Science*, pages 428–441. Springer, 2004. 1
- [2] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio, and K. R. Mueller. The Berlin brain-computer interface: Non-medical uses of BCI technology. *Frontiers in Neuroscience*, 4(0), 2010. 5
- [3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 2011. 5
- [4] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman. Geodesic star convexity for interactive image segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3129–3136, June 2010. 1
- [5] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum. Lazy snapping. *ACM SIGGRAPH*, 23(3):303–308, 2004. 1
- [6] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 22(140):1–55, 1932. 4
- [7] D. Liu, K. Pulli, L. G. Shapiro, and Y. Xiong. Fast interactive image segmentation by discriminative clustering. In *Proceedings of the 2010 ACM multimedia workshop, MCMC*, pages 47–52, 2010. 1
- [8] R. Lowry. Concepts & applications of inferential statistics. vassar college, poughkeepsie, ny, 2010. 3
- [9] G. H. M. Sadeghi, G. Tien and M. S. Atkins. Hands-free interactive image segmentation using eyegaze. SPIE Medical Imaging Conference, 2009. 1
- [10] 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. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001. 2, 3
- [11] K. McGuinness and N. E. O'Connor. A comparative evaluation of interactive segmentation algorithms. *Pattern Recogn.*, 43(2):434–444, Feb. 2010. 1, 4
- [12] P. Perez, A. Blake, and M. Gangnet. JetStream: probabilistic contour extraction with particles. *Proc. IEEE Int Conf. Computer Vision, ICCV'01*, 2001. 1
- [13] H. Petrie and N. Bevan. The evaluation of accessibility, usability and user experience. In C. Stephanidis, editor, *The Universal Access Handbook*. CRC Press. 1
- [14] C. Rother, V. Kolmogorov, and A. Blake. GrabCut: interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH*, pages 309–314, 2004. 1
- [15] T. Vallin Spina, A. Falco, and P. Vechiatto Miranda. User-steered image segmentation using live markers. volume 6854 of *CAIP*, pages 211–218. Springer, 2011. 2

7. Acknowledgments

We would like to thank Science Foundation Ireland grant 07/CE/11147 for funding this research. Many thanks to Prof. H. Lee and Dr. R. Batal for useful feedback on design. We are thankful to volunteers who conducted experiments.