Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2011 ; 2011: 1645–1648. doi:10.1109/ISBI.2011.5872719.

ACTIVE LEARNING GUIDED INTERACTIONS FOR CONSISTENT IMAGE SEGMENTATION WITH REDUCED USER INTERACTIONS

Harini Veeraraghavan and James V. Miller

General Electric Research, 1 Research Circle, Niskayuna, NY, USA.

Abstract

Interactive techniques leverage the expert knowledge of users to produce accurate image segmentations. However, the segmentation accuracy varies with the users. Additionally, users may also require training with the algorithm and its exposed parameters to obtain the best segmentation with minimal effort. Our work combines active learning with interactive segmentation and (i) achieves as good accuracy compared to a fully user guided segmentation but with significantly lower number of user interactions (on average 50%), and (ii) achieves robust segmentation by reducing segmentation variability with user inputs. Our approach interacts with user to suggest gestures or seed point placements. We present extensive experimental evaluation of our results on two different publicly available datasets.

Keywords

Active learning; SVM classification; interactive segmentation; learning based user guidance

1. INTRODUCTION

Accurate medical image segmentation is important for applications including computer aided diagnosis, therapy planning, and treatment. Compared to automatic approaches, interactive techniques [1, 2, 3, 4, 5] can produce more accurate segmentations, albeit with more user inputs and larger variability in the accuracies. To be viable for practical applications, an interactive approach must (a) minimize user interaction (b) minimize segmentation variability with users and (c) be computationally fast to allow fast user editing. Our work addresses (a),(b), and (c) by augmenting an interactive segmentation algorithm (specifically the GrowCut) [4] with support vector machine (SVM)-based active learning. As opposed to typical one-way interaction for segmentation, in our approach, the algorithm interacts with the user and suggests the placement of gestures. To handle image noise, correlated pixels, and the computational cost of selecting from $n \times m$ pixels, we employ a two-phase approach for gesture suggestion. First, the algorithm extracts query candidate pixels by combining segmentations produced by the GrowCut and SVM classification. Second, using SVM margin-based criteria [6] gestures are selected from the query candidate pixels. The segmentation improves iteratively with every suggestion. Fig. 1 shows the segmentations using the GrowCut with each algorithm suggestion accepted and labelled by the user.

Besides segmentation, our approach learns a model of the segmented target with much fewer number of labelled examples than fully supervised techniques [7, 8]. Our approach does not require a user to label whole images [9], and is not restricted to classifying discrete data [6, 10]. Unlike [11] which employs an iterative probabilistic framework for segmenting aligned images, our approach does not require the novel images to be aligned with the training image. As the user can modify their interaction, our approach can recover from local minima in learning resulting from the placement of starting gestures. Our approach uses a similar two-way interaction as [12, 13], but learns from a single image to segment medical images which contain much less texture and color compared to natural images. To our knowledge, ours is the first to employ active learning with an interactive segmentation for segmenting medical images.

2. ACTIVE LEARNING

Active learning is an iterative machine learning approach that models the data using a small number of labelled training examples by proactively selecting specific unlabelled examples for labelling. The goal of learning is to learn the best model of the data as fast (using as few examples) as possible. Successful learning is achieved by selecting the most *informative* example(s) for querying in each iteration. The most informative example is usually the one that is most difficult to classify [14, 6]. Our work is inspired by [6] which employs support vector machines (SVM) margins for example selection. The *margin* of a SVM is the distance of the closest training data of either class from the classification hyperplane. The *support vectors* are the training examples on the margin.

To formalize, given examples $\{x_1, \dots, x_n\}$ which are vectors from some d dimensional space $X \subseteq \mathbb{R}^d$ and their corresponding labels $\{y_1, \dots, y_n\}$, where $y \in \{-1, 1\}$ the SVM maps the original data into a high-dimensional space using a kernel function K as:

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) \quad (1)$$

where the kernel operator $K(u, v)$ can be expressed as an inner product $K(u, v) = \Phi(u) \cdot \Phi(v)$, simplifying the above equation to $f(x) = (\sum_{i=1}^n \alpha_i \Phi(x_i))$ [14]. The α_i are nonzero for the support vectors. Adding a new example to the training set of the SVM will either: (i) leave the margin unchanged, meaning the new data adds no additional information, or (ii) increase the margin, meaning the new data helps to separate the classes better, or (iii) decrease the margin, meaning the new data introduces more ambiguity. Selecting the examples that fall in category (iii) for querying would therefore help to reduce the ambiguity in the classification. This scheme is called the *Simple Margin* in [6]. Using the above intuition, [6] proposed *MaxMin*, *MaxRatio* and *Hybrid* margins for query selection.

Let m_i^+ and m_i^- be the margins of SVMs S_i^+ and S_i^- obtained by adding the example x_i (with removal) with label 1 and -1 to the existing training set, respectively. The MaxMin margin chooses an example x_j with the maximum of $\min(m^+, m^-)$ of all the unlabelled examples.

The MaxRatio margin chooses an example x_i whose $\min\left(\frac{m_i^+}{m_i^-}, \frac{m_i^-}{m_i^+}\right)$ is the largest. The Hybrid margin switches between the two.

3. METHOD: ACTIVE LEARNING COMBINED INTERACTIVE SEGMENTATION

Fig. 1 summarizes our segmentation approach. User initializes the algorithm with gestures (red for background, and green for foreground) Fig. 1(1), that results in a GrowCut segmentation Fig. 1(2) and SVM classification. The algorithm produces gesture suggestions which are accepted and labelled by the user in Fig. 1(3). The newly labelled gestures are combined with all the previously labelled gestures to produce new segmentation Fig. 1(4) followed by gesture suggestion Fig. 1(5) until convergence in Fig. 1(6).

Naive application of active learning such as in Section 2 to our problem is difficult due to (a) the computational cost of training $n \times m \times 2$ SVMs for query selection in every iteration, (b) noise in the pixels, and (c) handling pixel correlations as each unlabelled pixel is treated as a sample drawn from i.i.d. To obviate the afore-mentioned difficulties, we employ a two-phase approach for gesture suggestion. The algorithm for gesture suggestion is depicted in Algorithm 1.

Algorithm 1:

Active Learning Combined Interactive Segmentation

Data: Image I , Labelled gestures $X_t^G = \{x_i^g, y_i^g\}$
Result: Segmented Image X^S , SVM model μ

- 1 Grow Cut Segment $X_t^S \leftarrow \{I, X_t^G\}$
- 2 Update Learning $\mu_t \leftarrow \{F : f(X_t^G)\}$
- 3 Classify Image $X_t^L \leftarrow \{I, \mu_t\}$
- 4 Get Contradiction Image $X_t^C \leftarrow X_t^S \otimes X_t^L$
- 5 **if** $X_t^C = \emptyset$ **then**
- 6 **return** $(X^S \leftarrow X_t^S, \mu \leftarrow \mu_t)$
- 7 **else**
- 8 Select Query: $q_{t+1} \leftarrow \{M, X_t^C\}$
- 9 Add labels $X_{t+1}^G \leftarrow X_t^G \cup L_{t+1}$ and **go to** 1
- 10 **end**

First, we extract a small set of candidate query pixels by treating the SVM classification and the GrowCut segmentation as a diverse ensemble like [15]. The candidate pixels are analysed in the second phase using the SVM margin criteria (Section 2) to produce gesture suggestions. The candidate queries are pixels whose label assignments in the ensemble disagree. The GrowCut segmentation [4] is a competitive region growing segmentation using

the principles of cellular automata. The SVM classifier is trained using the intensities and Gabor features of the labelled gestures.

The gestures from the current iteration $X_t^G = \{\langle x_i, y_i^g \rangle\}$ produce two segmentations; GrowCut $X_t^S = \{\langle x_i, y_i^S \rangle | i = 1, \dots, N\}$ (Line 1 Algorithm 1), and current SVM classification $X_t^L = \{\langle x_i, y_i^L \rangle | i = 1, \dots, N\}$ (Line 3 Algorithm 1). An example user input and the corresponding GrowCut and SVM classification for the same input are depicted in Fig. 2(a), (b), and (c). The two segmentations X_t^S and X_t^L are combined to extract a contradiction label image $X_t^C = \{\langle x_i, y_i^S \neq y_i^L \rangle | i = 1, \dots, N\}$ (Line 4 Algorithm 1) also shown in Fig. 2(d). Next, using one of the SVM margin criteria, namely, the MaxMin, the MaxRatio, or the Hybrid margin as explained in the previous Section 2, a query pixel is selected (Line 8 Algorithm 1), also depicted in *cyan* in Fig. 2(e). The user labelled pixels from the current iteration L_{t+1} are added to the labelled gestures X_{t+1}^G (Line 9 Algorithm 1) and the algorithm is repeated until convergence. The GrowCut segmentation using the newly added gestures in Fig. 2(e) is shown in Fig. 2(f).

The algorithm stops when the classification and the segmentation are alike, i.e., $X_t^C = \emptyset$.

Our algorithm has a clear stopping condition (always terminates), and can be combined with other segmentation approaches. The user is not restricted to follow the suggested gestures and may paint where the deem appropriate.

4. EXPERIMENTAL EVALUATION AND RESULTS

Our experiments evaluated (a) consistency of segmentation using the algorithm generated gesture suggestions, and (b) reduction in the user interactions (as the length of gestures) using active learned model priors. We used the SPL tumor data sets [16] and the OASIS Alzheimer's database for segmenting ventricles. To eliminate any bias due to training of the user with the segmentation algorithm, all of our experiments are bootstrapped using computer generated gestures produced and labelled using the groundtruth as the **oracle**.

Fig. 3 depicts the results of a paired t-test that compared the segmentation accuracy using the initial labels produced from the **oracle** and the segmentation accuracy with algorithm suggested gestures upto a maximum of 5 iterations. The addition of the algorithm suggested gestures produce a significant difference in the segmentation accuracy, measured as DICE overlap scores with the groundtruth. As shown in Fig. 3, the overlap scores increase with iterations (indicated by increasing negative t-values). Fig. 4 shows some examples of segmentation using the initial labels alone and at the end of algorithm suggestions accepted and labelled using the oracle. The location of the initial labels are magnified for I, II, and IV. As shown, the segmentation resulting from accepting the gesture suggestions shown in Fig. 4(c) are more accurate than using the initial inputs alone in Fig. 4(b).

Fig. 5 shows a comparison of the accuracies in segmentation obtained by employing the active learned models as priors (I) and without any learning (II). In the latter case (II), a

human user guided the segmentation with as many inputs as required to produce the best segmentation, whereas in the former case with learned priors (I), the interactive segmentation was terminated when the segmentation accuracy was close to (II). Fig. 5(a) shows the accuracies for each of the active learning margins compared to the basic grow cut segmentation of a few exemplars selected from Fig. 5(b). Fig. 5(b) shows the relative difference in the accuracies between (I) and (II). As shown, the variation in the accuracies is -10% , 40% . In other words, the segmentations using learning was utmost 10% worse than the fully user guided segmentation without learning. The average accuracies of the segmentation using learning was 83% and with grow cut with no learning was 81% . Fig. 5(c) shows the number of gestures required for attaining the accuracies depicted in Fig. 5(b) with (I) and without learning (II). As shown, the number of gestures required for the segmentation using learning are much lower than when not using any learning. On an average, the number of gestures required in the case (I) with learning were 50% lower than in case (II) without learning. One limitation of the approach is that really poor placement of gestures to bootstrap our algorithm can result in suggestions being placed in seemingly irrelevant locations. As an interesting sideeffect of learning, incorrect user labels tend to make the algorithm repeatedly suggest queries around the areas of incorrect labellings. This we believe renders the algorithm robust to user errors.

5. CONCLUSIONS

In this work, we presented an approach for interactive segmentation that combines active learning with the GrowCut interactive segmentation. Using a two-way interaction approach our algorithm suggests locations for drawing gestures to the user, who in turn can label the pixels as suggested or pick where to draw. We showed that using active learning guided gesture suggestions reduces the variability of the segmentation and reduces the user interactions by almost (50%) compared to segmenting the novel images with no learning. Additionally, the learning is completely transparent to the user and does not require the user to explicitly provide a lot of labelled data for learning. Our approach is not restricted to the Grow-Cut segmentation and can be combined with any interactive segmentation algorithm.

Acknowledgments

This work was supported in part by the NIH NCRR NAC P41-RR13218 and is part of the National Alliance for Medical Image Computing (NAMIC) funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. The OASIS datasets were made available thanks to grants P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584.

References

- [1]. Boykov Y. and Jolly M-P, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in IEEE ICCV, 2001, pp. 105–112.
- [2]. Grady L, "Random walks for image segmentation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1768–1783, 2006.
- [3]. Letteboer MMJ, Olsen OF, Dam EB, Willems PWA, Viergever MA, and Niessen WJ, "Segmentation of brain tumors in magnetic resonance brain images using an interactive multiscale watershed algorithm," Academic Radiology, vol. 11, no. 10, pp. 1125–1138, 2004. [PubMed: 15530805]

- [4]. Vezhnevets V and Konouchine V, "GrowCut - Interactive multi-label N-D image segmentation," in Proc. Graphicon, 2005, pp. 150–156.
- [5]. Mishra A, Wong A, Zhang W, Clausi D, and Feiguth P, "Improved interactive medical image segmentation using enhanced intelligent scissors (eis)," in IEEE Intl. Conf. Engineering in Medicine and Biology, 2008, pp. 3083–3086.
- [6]. Tong S and Koller D, "Support Vector Machine active learning with applications to text classification," Journal of Machine Learning Research, pp. 45–66, 2001.
- [7]. Etyngier P, Ségonne F, and Keriven R, "Active contour-based image segmentation using machine learning techniques," in MICCAI, 2007, pp. 891–899.
- [8]. Gerber S, Tasdizen T, Joshi S, and Whitaker R, "On the manifold structure of the space of brain images," in MICCAI, 2009.
- [9]. Farhangfar A, Greiner R, and Szepesvári C, "Learning to segment from a few well-selected training images," in ICML, 2009, pp. 305–312.
- [10]. Hoi S, Jin R, Zhu J, and Lyu M, "Batch mode active learning and its application to medical image classification," in ICML, 2006.
- [11]. Raviv TR, Van Leemput K, Menze BH, Wells WM, III, and Golland P, "Segmentation of image ensembles via latent atlases," Medical Image Analysis, vol. 14, pp. 654–665, 2010. [PubMed: 20580305]
- [12]. Xia T, Wu Q, Chen C, and Yu Y, "Lazy texture selection based on active learning," The Visual Computer, vol. 26, no. 3, pp. 157–169, 2009.
- [13]. Batra D, Parikh D, Jeibo L, and Chen T, "iCoseg: Interactive co-segmentation with intelligent scribble guidance," in IEEE CVPR, 2010, pp. 3169–3176.
- [14]. Campbell C, Cristianni N, and Smola A, "Query learning by large margin classifiers," in ICML, 2000.
- [15]. Melville P and Mooney RJ, "Diverse ensembles for active learning," in ICML, 2004, pp. 584–591.
- [16]. Kaus M, Warfield SK, Nabavi A, Black PM, Jolesz FA, and Kikinis R, "Automated segmentation of MRI of brain tumors," Radiology, vol. 218, no. 2, pp. 586–591, 2 2001. [PubMed: 11161183]

Fig. 1.

Lesion segmentation using initial user inputs followed by user labels accepted and placed on algorithm suggested locations.

Fig. 2.
Steps in query selection.

Iterations	1	2	3	4	5
p-value <	0.006	0.001	0	0	0
p-value $\neq$	0.011	0.002	0.001	0	0
t-value	-2.71	-3.42	-3.77	-4.72	-5.13

Fig. 3.

Analysis of paired t-tests comparing the DICE overlap scores using no suggestions with 1 to 5 iterations of suggestions.

Fig. 4.

Example segmentation results using active learning combined interactive segmentation.
 Number of gesture suggestion iterations for image (by row) I - 10, II - 6, III - 6, IV - 8

Fig. 5.
Segmentation accuracies and the number of gestures on novel images using active learned priors and basic grow cut.