

Multi-Modality Feature Transform: An Interactive Image Segmentation Approach

Moustafa Meshry
moustafa.meshry@alexu.edu.eg

Ahmed Taha
ahmed.taha@alexu.edu.eg

Marwan Torki
<http://www.eng.alexu.edu.eg/~mtorki/>

CSE Dept.
Alexandria University
Alexandria, Egypt

In this paper, we tackle the interactive image segmentation problem. Unlike the regular image segmentation problem, the user provides additional constraints that guide the segmentation process. In some algorithms, like [1, 4], the user provides scribbles on foreground/background (Fg/Bg) regions. In other algorithms, like [6, 8], the user is required to provide a bounding box or an enclosing contour to surround the Fg object, other outside pixels are constrained to be Bg. In our problem, we consider scribbles as the form of user-provided annotation.

Introducing suitable features in the scribble-based Fg/Bg segmentation problem is crucial. In many cases, the object of interest has different regions with different color modalities. The same applies to a non-uniform background. Fg/Bg color modalities can even overlap when the appearance is solely modeled using color spaces like RGB or Lab. Therefore, in this paper, we purposefully discriminate Fg scribbles from Bg scribbles for a better representation. This is achieved by learning a discriminative embedding space from user-provided scribbles. The transformation between the original features and the embedded features is calculated. This transformation is used to project unlabeled features onto the same embedding space. The transformed features are then used in a supervised classification manner to solve the Fg/Bg segmentation problem. We further refine the results using a self-learning strategy, by expanding scribbles and recomputing the embedding and transformations.

Figure 1 illustrates the motivation for this paper. Color features usually cannot capture different modalities available in the scribbles and successfully distinguish Fg from Bg at the same time. As we can see in figure 1(b), the RGB color space will eventually mix Fg/Bg scribbles. On the other hand, figure 1(c) shows that a well-defined embedding space can clearly distinguish between Fg and Bg scribbles, while preserving different color modalities within each scribble.

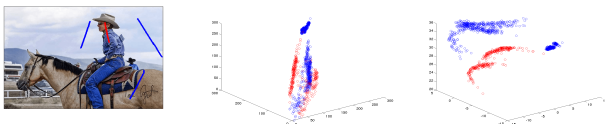


Figure 1: The effect of discriminative embedding. **Left (a)**: Image with provided user scribbles; red for Fg and blue for Bg. **Middle (b)**: 3D plot of the RGB channels for the provided scribbles. The scribbles are mixed in the RGB color space. **Right (c)**: 3D plot of the first 3 dimensions of our discriminative embedding. Color modalities present in the scribbles are preserved. Note that the Fg has two modalities, namely skin color and jeans. Also, the Bg has two modalities: the sky and horse body.

Our contributions in this paper are multifold; **First**, we present a novel representation of image features in the scribble-based Fg/Bg segmentation problem. **Second**, we utilize this representation in two novel interactive segmentation algorithms: (i) One-pass supervised algorithm, which we extend to (ii) a self-learning semi-supervised algorithm. **Third**, we present an extensive evaluation on a standard dataset with clear improvements over state-of-the-art algorithms.

The proposed segmentation algorithm learns a discriminative embedding space for the scribbles using a supervised dimensionality reduction technique, like LDA [2, 3] or LFDA [7]. LDA seeks to maximize the between-class separation while minimizing the within-class proximity. LFDA extends LDA by preserving the locality of features that belong to the same class. This is illustrated in figure 1, where Fg has two modalities (skin color and jeans) and Bg also has two modalities (sky and horse body). We then use the learned transformation matrix to transform pixels' color features by projecting them onto the new embedding space. Finally, we classify every pixel as Fg or Bg based on its embedding coordinates. To enhance the classification, we use an iterative version which

Transformation Method	Jaccard Index
No transformation	0.549 ± 0.260
1-pass LDA	0.627 ± 0.179
Iterative LDA	0.636 ± 0.180
1-pass LFDA	0.664 ± 0.184
Iterative LFDA	0.678 ± 0.180

Table 1: Segmentation results on ISEG dataset

expands the original scribbles and recomputes the whole pipeline until a stopping criterion is met. A final post processing step is used to remove small islands as done in [5]. Our methods are proved to outperform state-of-the-art algorithms on the standard ISEG dataset [4]. Table 1 shows the segmentation result of different feature embeddings. It is clear that careful embedding can elevate the results significantly. Figure 2 shows qualitative results of our approach.

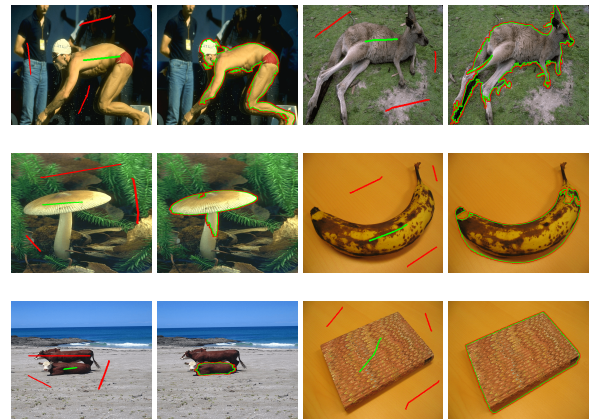


Figure 2: Qualitative results for 6 out of 151 images. First and third columns show the original image with user scribble annotation. Second and fourth columns show our output.

- [1] Yuri Y Boykov and M-P Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In *IEEE International Conference on Computer Vision, ICCV*, 2001.
- [2] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.
- [3] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic press, 2013.
- [4] Varun Gulshan, Carsten Rother, Antonio Criminisi, Andrew Blake, and Andrew Zisserman. Geodesic star convexity for interactive image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [5] Jiangyu Liu, Jian Sun, and Heung-Yeung Shum. Paint selection. In *ACM Transactions on Graphics (ToG)*, 2009.
- [6] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*, 2004.
- [7] Masashi Sugiyama. Local fisher discriminant analysis for supervised dimensionality reduction. In *Proceedings of the 23rd international conference on Machine learning*, pages 905–912. ACM, 2006.
- [8] Jiajun Wu, Yibiao Zhao, Jun-Yan Zhu, Siwei Luo, and Zhuowen Tu. Milcut: A sweeping line multiple instance learning paradigm for interactive image segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, 2014.