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

Abstract

Introducing suitable features in the scribble-based foreground-background (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. 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 the 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. Finally, we evaluate our algorithms and compare their performance against the state-of-the-art methods on the ISEG dataset with clear improvements over competing methods.

1 Introduction

Interactive image segmentation problem has gained a lot of interest from computer vision researchers. Unlike the regular image segmentation problem, the user provides additional constraints that guide the segmentation process. In some algorithms, like [3, 9, 18, 20], the user provides scribbles on Fg/Bg regions. In other algorithms, like [13, 17, 22], 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. A recent study on the different forms of user guidance to the interactive image segmentation problem can be found in [11]. In our problem, we consider scribbles as the form of user provided annotation.

The problem formulation of interactive segmentation benefits from the labels. This leads to two generic categories of algorithms, namely region growing like [1, 15] and graph-based solutions like [3, 14, 17, 20].

In figure 1 we illustrate the motivation for this paper. Color features usually cannot capture different modalities available in the scribbles and successfully distinguish Fg from
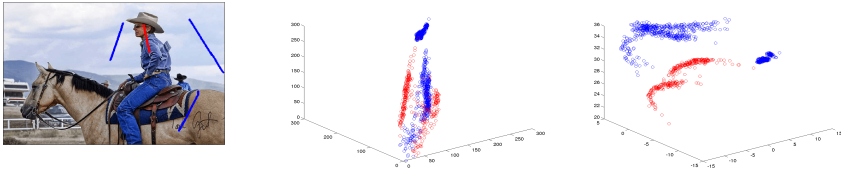
Figure 1: The effect of discriminative embedding. **Left (a)**: Image with provided user scribbles; red for Fg and blue for Bg. **Middle (b)**: 3D plot of the RGB channels for the provided scribbles. The scribbles are mixed in the RGB color space. **Right (c)**: 3D plot of the first 3 dimensions of our 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.

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.

In this paper, we present an approach for representing Fg/Bg scribbles in a way that combines the *appearance* and *class label disambiguation*. We first extract pixels' color features using standard color spaces, namely RGB and Lab. Next, we weight pixels' features by the geodesic distance to the nearest Fg scribble and Bg scribble respectively. Later, we learn a discriminative embedding space for the scribbles using a supervised dimensionality reduction technique, like LDA [7, 8] or LFDA [19]. We transform pixels' 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 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 [14]. Our methods are proved to outperform state-of-the-art algorithms on the standard ISEG dataset [9].

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 in section 3.3, which we extend to *(ii)* a self-learning semi-supervised algorithm in section 3.4. **Third**, we present an extensive evaluation on a standard dataset with clear improvements over state-of-the-art algorithms.

2 Related Work

Color Features: Many algorithms in the literature base their solution for Fg/Bg segmentation on color features. From the region growing family, the seeded region growing algorithm [1] iterates to assign a pixel to its nearest labeled point based on color distance. In MSRM [15], color histograms are built on top of pre-computed superpixels and the unlabeled regions are merged to similar labeled regions using the Bhattacharyya coefficient as a similarity measure. Color features are also utilized in the graph cut family. For example, in GrabCut [17] multiple color Gaussian Mixture Models are introduced for both the foreground and the background. An iterative procedure that alternates between estimation and parameter learning is utilized.

Geodesic Distance: Using Geodesic distance proved to be useful in interactive image segmentation. Many approaches like [2, 5, 9] define the geodesic distance between two points, $d_g(a, b)$, to be the smallest integral of a weight function over all possible paths from a to b. Hence, unlike Euclidean paths, geodesic paths can bend and take any arbitrary shape. This property is suitable for image segmentation since the boundaries of an object can take any arbitrary shape as well. We follow the same notation presented in [9], which first defines the length of a discrete path $L(\Gamma)$ as:

$$L(\Gamma) = \sum_{i=1}^{n-1} \sqrt{(1 - \gamma_g) d(\Gamma^i, \Gamma^{i+1})^2 + \gamma_g \| \nabla I(\Gamma^i) \|^2} \qquad (1)$$

where Γ is an arbitrary parametrized discrete path with n pixels given by $\Gamma^1, \Gamma^2, \cdots, \Gamma^n$, $d(\Gamma^i, \Gamma^{i+1})$ is the Euclidean distance between successive pixels, and the quantity $\| \nabla I(\Gamma^i) \|^2$ is a finite difference approximation of the image gradient between the points (Γ^i, Γ^{i+1}). The parameter γ_g weights the Euclidean distance with the geodesic length. The geodesic distance between two points, $d_g(a, b)$, is then defined as:

$$d_g(a, b) = \min_{\Gamma \in P_{a,b}} L(\Gamma) \quad , \quad \Gamma_{a,b} = arg \min_{\Gamma \in P_{a,b}} L(\Gamma) \qquad (2)$$

where $P_{a,b}$ denotes the set of all discrete paths between two grid points a and b. The above definition of geodesic distance between two points extends to distance between a set of points c, namely user scribbles, and a point p exactly as in [5].

Linear Dimensionality Reduction: Methods of linear dimensionality reduction are used to produce a lower dimensional representation of the feature points. For example the Principal Component Analysis(PCA) [12] tries to minimize the after projection reconstruction error by projecting the data on the directions of larger variance. Locality Preserving Projections (LPP) [10] tries to embed the close features in the original feature space into close embedded features. Despite the effectiveness of PCA and LPP as unsupervised dimensionality reduction algorithms, both do not benefit from existing labels.

In many applications, the data is labeled or partially labeled and hence supervised methods for dimensionality reduction seek to maximize the between-class separation, while minimizing the within-class proximity, like linear discriminant analysis (LDA) [7, 8] and its variants [19, 23, 24, 25]. A special variant of LDA is the LFDA algorithm by Sugiyama [19]. In LFDA, the embedded coordinates will be like LPP, as it preserves the locality of features that belong to the same class while maximizing the between-class separability like LDA.

3 Approach

3.1 Supervised Dimensionality Reduction Using LDA and LFDA

Formulation: We follow the formulation of [19] for supervised dimensionality reduction. Let $x_i \in \Re^d (i = 1, 2, \cdots n)$ be d-dimensional samples and $y_i \in \{1, 2, \cdots c\}$ be the associated class labels, where n is the number of samples and c is the number of classes. Let n_i be the number of samples in class i. Let X be the data matrix: $X = (x_1|x_2|...|x_n)$. Let $z_i \in \Re^m (1 \le m \le d)$ be the embedded samples, where m is the dimensionality of the embedding space. We are seeking to find a $d \times m$ transformation matrix T, such that z_i is given by:

$$z_i = T^T x_i \qquad (3)$$

Linear Discriminant Analysis (LDA): The objective of LDA embedding is to find a transformation matrix T. The matrix T best transforms the original features X into new coordinates Z to discriminate embedded features of predefined classes. To obtain such a transformation matrix, a between-class scatter matrix, S_b, and a within-class scatter matrix, S_w, are defined as:

$$S_b = \sum_{k=1}^{c} n_k (\mu_k - \mu)(\mu_k - \mu)^T \quad , \quad S_w = \sum_{k=1}^{c} \sum_{i \in C_k} n_k (x_i - \mu_k)(x_i - \mu_k)^T \tag{4}$$

where μ_k is the sample mean of class C_k and μ is the sample mean of all the features. The transformation matrix T is then defined as:

$$T = \arg \max_{T \in \mathcal{R}^{d \times m}} tr((T^T S_w T)^{-1} T^T S_b T) \tag{5}$$

The matrix T is constructed by concatenating the generalized eigen vectors corresponding to the largest m generalized eigen values of the problem $S_b v = \lambda S_w v$

Local Fisher Discriminant Analysis (LFDA): The objective of LFDA is similar to that of LDA. It looks for a transformation matrix T_l. In addition to the separation between different classes, the transformation matrix T_l is sought to preserve the locality in each of the classes. This is done by defining the between-class scatter matrix, S_b', and the within-class scatter matrix, S_w', as:

$$S_b' = \frac{1}{2} \sum_{i,j=1}^{n} A_b(i,j)(x_i - x_j)(x_i - x_j)^T \quad , \quad S_w' = \frac{1}{2} \sum_{i,j=1}^{n} A_w(i,j)(x_i - x_j)(x_i - x_j)^T \tag{6}$$

where

$$A_b(i,j) = \begin{cases} W_{ij}(\frac{1}{n} - \frac{1}{n_c}) & \text{if } y_i = y_j = c, x_j \in N(x_i) \\ 0 & \text{otherwise} \end{cases}, A_w(i,j) = \begin{cases} \frac{W_{ij}}{n_c} & \text{if } y_i = y_j = c, x_j \in N(x_i) \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

W_{ij} represents the affinity between points x_i and x_j. The sample $x_j \in N(x_i)$ if x_j is one of the k-nearest neighbors of x_i or vise versa in the same class c. The transformation matrix T_l is defined as:

$$T_l = \arg \max_{T_l \in \mathcal{R}^{d \times m}} tr((T_l^T S_w' T_l)^{-1} T_l^T S_b' T_l) \tag{8}$$

Similarly, the matrix T_l is constructed by concatenating the generalized eigen vectors corresponding to the largest m generalized eigen values of the problem $S_b' v = \lambda S_w' v$.

3.2 Color-Geodesic Features

The first step in our approach to interactive image segmentation is to transform the original feature vector x_i of every pixel i into a new feature vector z_i in the embedding space. Thus, we start by defining our original space pixels' features x_i. For each pixel, we extract both the RGB and Lab channels. We normalize each of these 6 channels to be between $[0,1]$. Then, we aim to weight this color vector based on the pixels' geodesic distance to Fg scribbles, such that pixels geodesically close to Fg scribbles would have higher weight. This is achieved by, first, computing the geodesic distance of all pixels to Fg scribbles. Then, normalizing the distance to be between $[0,1]$. Finally, we define our Fg distance weighting of pixel p_i to
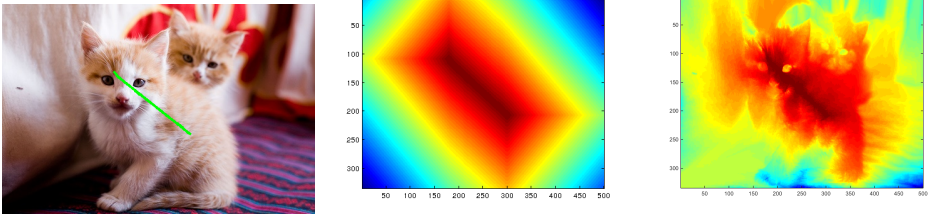
Figure 2: The effect of weighting color channels. **Left**: Original image with the Fg scribble **Middle**: Manhattan distance weights with respect to the Fg scribble; red means higher weight and blue means lower weight. **Right**: Geodesic distance weights with respect to the Fg scribble.

be $w_i = 1 - \text{normalized}(\text{GeoDist}(p_i, \text{Fg}))$. So, we obtain 6 color channels that represent RGB and Lab values, weighted by their geodesic distance to Fg scribbles. We apply the same procedure but for Bg scribbles, and we concatenate the Fg and Bg weighted color vectors to get our original space feature vector $x_i \in \Re^{12}$. Figure 2 shows the effect of weighting the RGB color channels using geodesic distance to the closest Fg scribble.

In the next two sections, we show how to use the defined feature vector x_i in our Multi-Modality Feature Transform approach for interactive image segmentation.

Algorithm 1 One-Pass MMFT Classification

INPUT: Image I, scribble masks FgMask and BgMask, K

OUTPUT: Labels vector y for every pixel

1: Compute the data matrix X according to section 3.2
2: Extract scribble data matrix X_{scr} and y_{scr} for $pixel_i \in$ FgMask or $pixel_i \in$ BgMask
3: Calculate the Transformation Matrix T_l using equation 8
4: Compute the transformed coordinates $Z = T_l^T X$
5: **for all** $z_i \in Z$ **do**
6:     find the nearest K features that have $y =$ Fg and compute $D_{\text{averageFg}}$
7:     find the nearest K features that have $y =$ Bg and compute $D_{\text{averageBg}}$
8:     compute $DistRatio_i = log(D_{\text{averageFg}}/D_{\text{averageBg}})$
9:     **if** $DistRatio_i < 0$ **then**
10:         $y_i =$ Fg
11:     **else**
12:         $y_i =$ Bg
13:     **end if**
14: **end for**
15: Apply post processing on y as in [14]

3.3 One-Pass Multi-Modality Feature Transform (O-MMFT)

In this section, we introduce our one-pass MMFT algorithm 1, which is a supervised Fg/Bg classification algorithm. It starts with feature extraction, as in 3.2, to compute the $d \times n$ data matrix X, where d is the dimensionality of pixels' feature vector, x_i, and n is the number of pixels. Next, we select labeled pixels $l = l_{\text{Fg}} \cup l_{\text{Bg}}$ to define a $d \times l$ scribble data matrix X_{scr}. We apply a supervised dimensionality reduction algorithm to obtain the transformation

matrix T_l using equation 8. We compute the transformed embedding coordinates for the data matrix X to find Z. We then apply a Nearest Neighbor search strategy to classify unlabeled features, in the transformed coordinates, into Fg/Bg classes; we find the closest k labeled features from Fg and average the distance to these k features. We do the same with respect to Bg features. For each pixel, the smaller average distance will denote the class label.

3.4   Iterative Multi-modality Feature Transform (I-MMFT)

We proceed to introduce our second algorithm 2, which is a self-learning semi-supervised Fg/Bg classification algorithm. In self-learning [4], a classifier is initially constructed using a labeled set, then its accuracy is enhanced by adding more data from the unlabeled set. Our initial classifier in this case is the O-MMFT 1. To manage the stopping criteria, we add two constraints: first, we do not exceed five iterations; second, we do not allow the matrix X_{scr} to exceed a predefined *MaxProblemSize*, which we set to 8000 points. These two constraints are used to maintain real time performance. In figure 3, we show examples to illustrate the improvements of the Iterative MMFT algorithm over its one-pass counterpart.

Algorithm 2 Iterative MMFT Fg/Bg Classification

INPUT: Image I, FgMask, BgMask, K, MaxIter, MaxProblemSize
OUTPUT: Labels vector y for every pixel

1: Set $DistRatioOld_i = zero \forall i$; Set $Iter = 1$
2: **repeat**
3:     Compute the data matrix X according to section 3.2
4:     Extract scribble data matrix X_{scr} and y_{scr} for $pixel_i \in$ FgMask or $pixel_i \in$ BgMask
5:     Calculate the Transformation Matrix T_l using equation 8
6:     Compute the transformed coordinates $Z = T_l^T X$
7:     **for all** $z_i \in Z$ **do**
8:         find the nearest K features that have $y =$ Fg and compute $D_{averageFg}$
9:         find the nearest K features that have $y =$ Bg and compute $D_{averageBg}$
10:         compute $DistRatio_i = DistRatioOld_i + log(D_{averageFg}/D_{averageBg})$
11:         **if** $DistRatio_i < 0$ **then**
12:             $y_i =$ Fg
13:         **else**
14:             $y_i =$ Bg
15:         **end if**
16:     **end for**
17:     Update FgMask and BgMask with most confident features.
18:     $Iter = Iter + 1$; $ProblemSize = SizeFgMask + SizeBgMask$;
19:     $DistRatioOld = DistRatio$
20: **until** MaxIter is reached or MaxProblemSize is reached
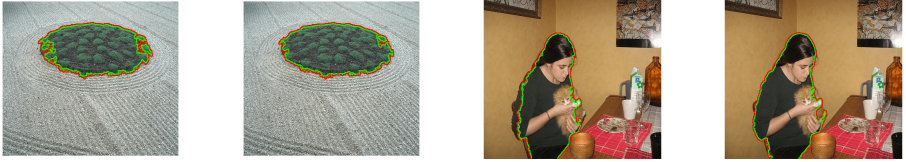21: Apply post processing on y as in [14]

Figure 3: Comparison between O-MMFT and I-MMFT. **Left and Middle-Right**: Result using O-MMFT. **Middle-Left and Right**: Result using I-MMFT.

4 Experiments

In this section, we show the effectiveness of our novel features and algorithms. We start by presenting two control experiments on the One-pass Multi-Modality Feature Transform (O-MMFT) algorithm 1. The first experiment demonstrates the effect of changing the transformation dimensionality m in the LFDA embedding step of our O-MMFT approach. The second experiment demonstrates the effect of the Geodesic weighting on color channels. Finally, we compare our algorithms to the state-of-the-art methods both quantitatively and qualitatively.

Dataset: To evaluate our Multi-Modality Feature Transform (MMFT) approach quantitatively, we use the Geodesic Star-Dataset [9]. The dataset is a well known scribble-based interactive image segmentation benchmark. It exhibits different variations, since it comes from a collection of different datasets. It consists of 151 images: 49 images taken from the GrabCut dataset [17], 99 from the PASCAL VOC dataset [6], and 3 images from the Alpha matting dataset [16]. For every image, the dataset simulates a user input in the form of an image containing four user scribbles. These four fixed scribbles are divided into one foreground and three background scribbles.

Evaluation Measure: We use Jaccard Index (overlap ratio) as a measure of accuracy of different segmentation algorithms. Jaccard Index is used to evaluate segmentation quality in the VOC segmentation challenge [6].

4.1 Control Experiments

Effect of dimensionality m in LFDA: One of the key differences between LDA and LFDA is the rank of the S_b matrix. In case of **LDA**, it is rank deficient and its rank is equal to the number of classes (two in our problem). However, this is not the same for LFDA as its S'_b matrix is not rank deficient and might be of full rank. This gives additional flexibility to using LFDA over LDA since it can produce richer dimensions. We use LFDA as a building block in our O-MMFT algorithm. Hence, we test the effect of changing the transformation dimensionality m. We observe in table 1 that the difference in Jaccard index is not significant and that ensures the stability of our algorithm with respect to the change of transformation dimensionality.

Effect of Geodesic Weighting: We choose different combinations of color features and distance weighting methods. Possible color features are RGB and Lab features. Possible

Figure 4: An example of the impact of Geodesic weighting on the O-MMFT segmentation result. **Left**: Image with provided user scribbles; green for Fg and red for Bg. **Middle**: Segmentation output in case of no Geodesic weighting on pixels' features. **Right**: Segmentation output in case of weighting pixels' features by their Geodesic distance to Fg and Bg scribbles.

LFDA Transformation Dimension	Jaccard Index
3	0.655
6	0.664
9	0.662
12	0.660

Table 1: Role of changing m in LFDA for our O-MMFT algorithm

distance weighting methods are the proposed procedure in section 3.2 and the basic distance transform using Manhattan distance.[1] Table 2 shows that best results are obtained via the procedure proposed in section 3.2. Incorporating geodesic weighting improved the segmentation accuracy from 0.413 to 0.664. Figure 4 shows an example of how Geodesic weighting affects the segmentation result in our approach. One-pass MMFT approach reaches best results with the following settings: 6 dimensions LFDA transformation, and geodesic weighting over Lab and RGB color models.

Features(weighting)	Jaccard Index
Lab(none)+RGB(none)	0.413
RGB(Manhattan)	0.632
Lab(Manhattan)	0.597
RGB(Geodesic)	0.654
Lab(Geodesic)	0.663
RGB(Geodesic) + Lab(Geodesic)	0.664

Table 2: Role of the geodesic weighting. Results are obtained using O-MMFT with $m = 6$

4.2 Comparative Evaluation

We compare our results[2] to state-of-the-art methods over the whole ISEG dataset. We present in table 3 different algorithms, namely Graph-Cut of Boykov and Jolly(BJ) [3], Post processing on BJ (PP) [14], Shortest Path (SP-SIG) [2], Geodesic Star Convexity (GSC) [9], Euclidean Star Convexity (ESC) [21] and Seeded Laplacian (SL) [20]. We show the effect of the feature transformation in comparison to the state-of-the-art methods. The transformation methods produce better results than competing algorithms. In addition, both our

---

[1]We use the MATLAB function bwdist on the binary Fg and Bg scribble mask images.
[2]The MATLAB source code is available on the author's webpage.

Method	Jaccard Index
BJ [3]	0.496 ± 0.261
PP [14]	0.593 ± 0.255
GSC [9]	0.615 ± 0.257
ESC [21]	0.610 ± 0.249
SP-SIG [2]	0.622 ± 0.171
SL [20]	0.670 ± 0.18

Transformation Method	Jaccard Index
No transformation	0.549 ± 0.260
O-LDA	0.627 ± 0.179
I-LDA	0.636 ± 0.180
O-MMFT	0.664 ± 0.184
I-MMFT	0.678 ± 0.180

Table 3: Quantitative comparison with state-of-the-art algorithms

O-MMFT and I-MMFT algorithms outperform the O-LDA and I-LDA. This result confirms our intuition about the importance of preserving the locality within each of the Fg/Bg classes.

In figure 5, we show a qualitative comparison between our iterative-MMFT approach and the state-of-the-art. The qualitative results reflect the improvements achieved using our novel approach.

5 Conclusion

In this paper, we presented a novel approach to define suitable features for the interactive image segmentation problem. To learn these features, we transformed the Geodesically weighted color features to an embedding space using a supervised dimensionality reduction technique. Several advantages were gained due to this transformation: First, features of different classes are separated from each other. Second, features from the same class with the same modality are transformed close to each other. Third, the transformation does not restrict features of the same class with different modalities to be close. These three advantages allowed us to use a simple classification procedure in the O-MMFT algorithm that we proposed. We enhanced the results using a self-learning strategy that we refer to as the I-MMFT algorithm. Our control experiments emphasized the effect of several factors like the Geodesic weighting and the dimensionality choice for our algorithms. Our comparative evaluation experiments showed that, using the MMFT algorithm, we can do better than state-of-the-art algorithms that use shape constraints [9, 21] while we do not.
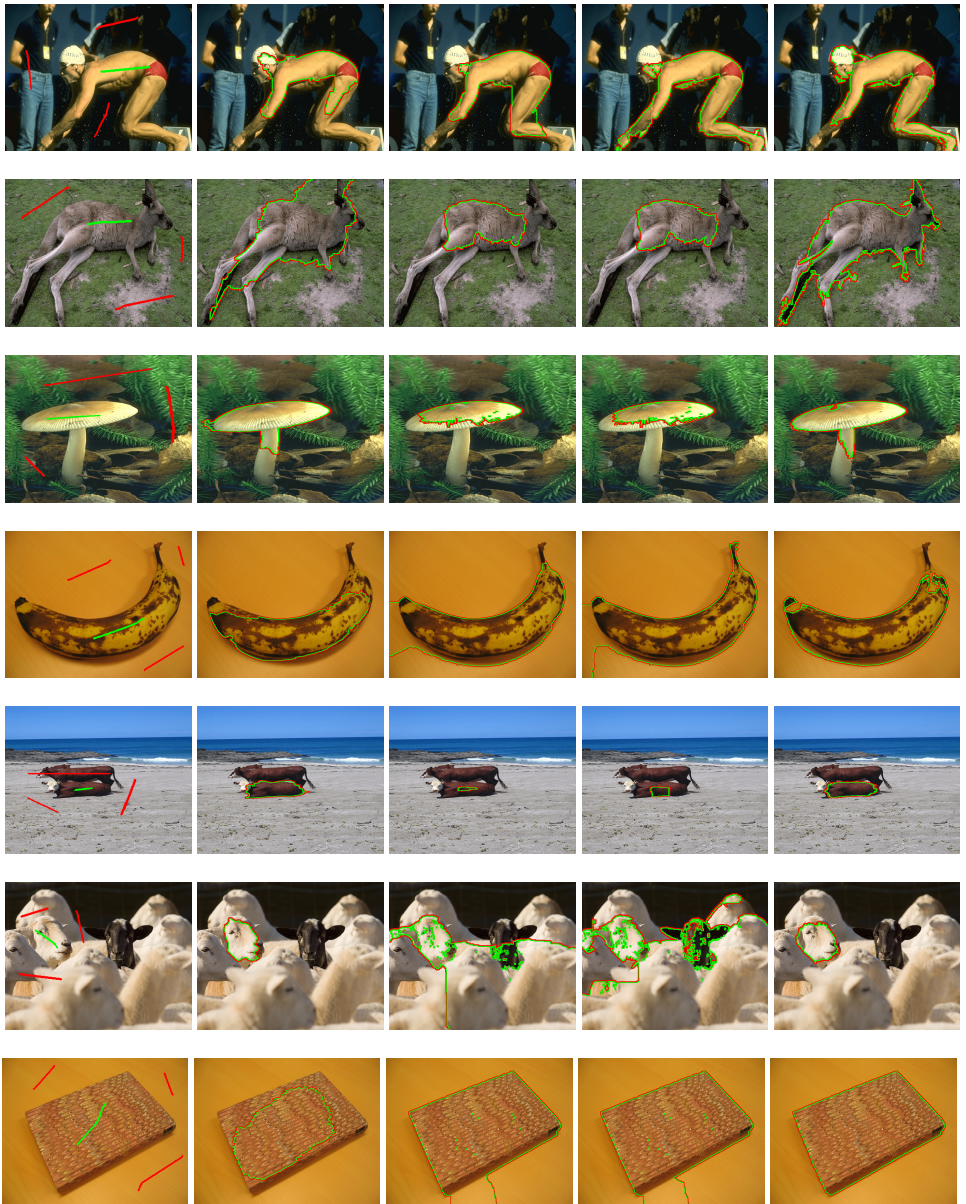
6 Acknowledgement

Figure 5: Qualitative results for 7 out of 151 images. The first column shows the original image with user scribble annotation. The rest of columns shows the segmentation results using SP-SIG, GSC, PP, **I-MMFT** methods respectively.

References

[1] Rolf Adams and Leanne Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994.

[2] Xue Bai and Guillermo Sapiro. A geodesic framework for fast interactive image and video segmentation and matting. In *IEEE 11th International Conference on Computer Vision, ICCV.*, 2007.

[3] 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.

[4] Olivier Chapelle, Bernhard Schölkopf, Alexander Zien, et al. *Semi-supervised learning*. MIT press Cambridge, 2006.

[5] Antonio Criminisi, Toby Sharp, and Andrew Blake. Geos: Geodesic image segmentation. In *Computer Vision–ECCV*, 2008.

[6] M Everingham, L Van Gool, C Williams, J Winn, and A Zisserman. The pascal visual object classes challenge 2009. In *2nd PASCAL Challenge Workshop*, 2009.

[7] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[8] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic press, 2013.

[9] 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.

[10] Xiaofei He and Partha Niyogi. Locality Preserving Projections. In *Advances in Neural Information Processing Systems 16*, 2004.

[11] Suyog Dutt Jain and Kristen Grauman. Predicting sufficient annotation strength for interactive foreground segmentation. In *IEEE International Conference on Computer Vision (ICCV).*, 2013.

[12] I. T. Jolliffe. *Principal component analysis*. Springer, second edition, October 2002. ISBN 0387954422.

[13] Victor Lempitsky, Pushmeet Kohli, Carsten Rother, and Toby Sharp. Image segmentation with a bounding box prior. In *IEEE 12th International Conference on Computer Vision*, 2009.

[14] Jiangyu Liu, Jian Sun, and Heung-Yeung Shum. Paint selection. In *ACM Transactions on Graphics (ToG)*, 2009.

[15] Jifeng Ning, Lei Zhang, David Zhang, and Chengke Wu. Interactive image segmentation by maximal similarity based region merging. *Pattern Recognition*, 2010.

[16] Christoph Rhemann, Carsten Rother, Jue Wang, Margrit Gelautz, Pushmeet Kohli, and Pamela Rott. A perceptually motivated online benchmark for image matting. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2009.

[17] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics (TOG)*, 2004.

[18] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.

[19] 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.

[20] Ahmed Taha and Marwan Torki. Seeded laplacian: An interactive image segmentation approach using eigenfunctions. In *Proc. IEEE International Conference on Image Processing (ICIP)*, 2015.

[21] Olga Veksler. Star shape prior for graph-cut image segmentation. In *Computer Vision–ECCV*, 2008.

[22] 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.

[23] Dong Xu, Shuicheng Yan, Dacheng Tao, Stephen Lin, and Hong-Jiang Zhang. Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval. *IEEE Transactions on Image Processing*, 2007.

[24] Jieping Ye and Tao Xiong. Computational and theoretical analysis of null space and orthogonal linear discriminant analysis. *Journal of Machine Learning Research*, 2006.

[25] Jieping Ye, Ravi Janardan, Qi Li, and Haesun Park. Feature reduction via generalized uncorrelated linear discriminant analysis. *IEEE Transactions on Knowledge and Data Engineering*, 2006.