

Interactive Image Segmentation Based on Gaussian Mixture Models with Spatial Prior

Mo Yan¹ and Peng-Lang Shui²

^{1,2}National Lab of Radar Signal Processing, Xidian University, Xi'an 710071,
China
xmualex@yeah.net

Abstract

In this paper, an interactive color natural image segmentation method is proposed. The method extends the Gaussian Mixture Model (GMM) by taking into account user markers as useful spatial prior. In this way, a distribution combining with color and spatial location is obtained. The distribution is incorporated in a Bayesian MAP approach. The approach is formalized as an iterative energy minimization problem. A direct global minimization technique based on variational method is employed to get binary solution. After each iteration, the largest connected region from foreground is used to update foreground GMMs and achieve more superior performance than original GrabCut method. Extensive experiments are performed on public benchmark datasets such as GrabCut benchmark, Berkeley segmentation database and Graz benchmark. The results show that the proposed method can extract the object region from the complex background accurately.

Keywords: image segmentation, GMM, spatial prior, convex optimization

1. Introduction

Image segmentation is an important and a fundamental problem in computer vision. Its purpose is to divide an image into nonoverlapping regions. Through this procedure, one can extract interested objects from a complex background. In general, due to the complex color and texture features in a natural image and the user's interests are different; the fully automatic segmentation of the objects from the background becomes very hard. Therefore, interactive segmentation algorithms are widely used. Over the last few years, there are a number of breakthroughs in image segmentation community such as level set methods [1-3], graph cuts [4-7], convex optimization techniques [3, 8-10], edge-grouping [11] and region-merge technique [12], etc. on.

Among those methods, graph cuts, level set methods and convex optimization techniques have been shown to be powerful techniques for segmentation problem. They convert an image segmentation problem to an energy minimization. The objective functional typically contains a data fidelity term and a regularization term. The data fidelity term based on user inputs is used to measure the conformity of the image data within the marked regions to statistical models. The regularization term is used to smooth segmentation boundaries. Graph cuts based methods are discrete approaches based on Markov Random Fields. Level set methods and convex optimization techniques are continuous approaches based on partial differential equations [13].

Recently, a popular method based upon the graph cuts algorithm called GrabCut extends graph cuts to color images and incomplete trimaps. It uses Gaussian Mixture Model (GMM) to make color distribution and iteratively estimates parameters to do graph cuts algorithm. By drawing a rectangle box around the desired foreground, one can extract foreground object easily. Despite using GMM to estimate color distribution is more accuracy, GrabCut algorithm often fails when the unknown pixels which belong to

background (Figure 1.a) take part in the estimation of GMM parameters. In this case, since some background pixels are calculated to estimate GMM parameters from the first time, it makes the final segmentation failure (Figure 1.b). Another case is, when background has overlapping color distribution with foreground (Figure 2.b), algorithm fails.

In order to obtain a good segmentation, the objective functional should has a good data fidelity term which can assign pixels good values so that these pixels can fit corresponding regions very well. However, since the color distributions of foreground and background often overlap, the segmentation procedure often fails. Thus only considering color distribution just like GrabCut may lead to failure in dealing with some images. Inspired by spatially varying color distributions segmentation [2], this paper proposed a new GMM-based segmentation method. In this method, spatial location of user markers is integrated into GMM. Since connectivity is important for object extracting, only the largest foreground connected region after each rough segmentation is used in GMM learning to update parameters. Then the color distribution with spatial location is incorporated in a Bayesian MAP framework. The problem is converted to solve an energy minimization. A direct global minimization convex segmentation technique in [3] is employed to solve the problem. In this paper, we just consider a two-phase segmentation problem.

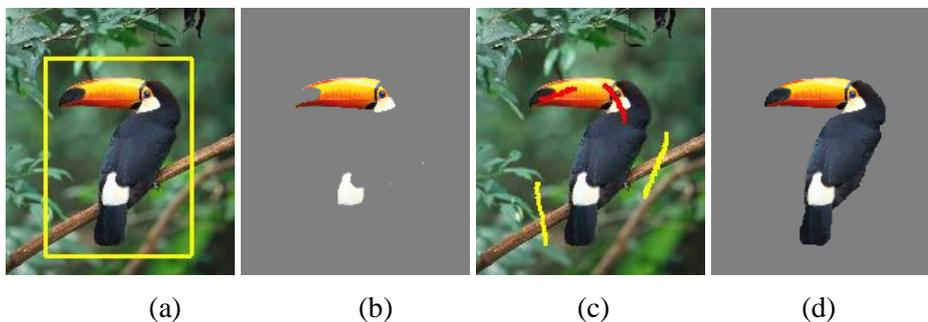


Figure 1. (a) GrabCut User Input (b) Segmentation Result by GrabCut (c) The Interactive Input by the User. The Red Lines are the Object Markers and the Yellow Lines are the Background Markers. (d) Segmentation Result by the Proposed Method

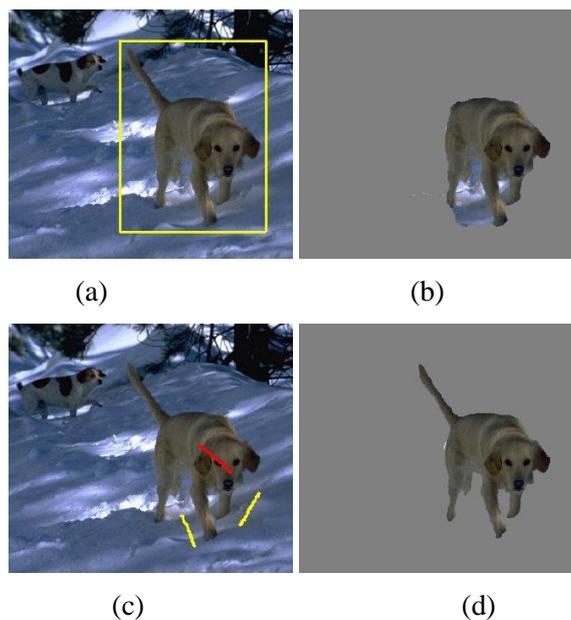


Figure 2. (a) GrabCut User Input (b) Segmentation Result by GrabCut (c) The Interactive Input by the User. The Red Lines are the Object Markers and the Yellow Lines are the Background Markers. (d) Segmentation Result by the Proposed Method

2. Gaussian Mixture Model (GMM)

A GMM is a parametric probability density function represented as a weighted sum of M Gaussian density functions. It is given by the equation,

$$p(\mathbf{x} | \mathbf{Q}) = \sum_{i=1}^M p_i g(\mathbf{x} | \mathbf{q}_i), \quad (1)$$

Where \mathbf{x} is a D -dimensional data vector, $p_i, i=1, L, M$ are the mixture weighting coefficients that satisfy $\sum_{i=1}^M p_i = 1$, and $g(\mathbf{x} | \mathbf{q}_i), i=1, L, M$ are called the Gaussian components. Each component is a D -variate Gaussian density function of the form,

$$g(\mathbf{x} | \mathbf{q}_i) = \frac{1}{(2\pi)^{D/2} (\det \mathbf{S}_i)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^T \mathbf{S}_i^{-1}(\mathbf{x} - \mathbf{m}_i)\right\}, \quad (2)$$

Where the parameter $\mathbf{q}_i = \{\mathbf{m}_i, \mathbf{S}_i\}$ is consist of mean vector \mathbf{m}_i and covariance matrix \mathbf{S}_i . These parameters are represented together by the notation $\mathbf{Q} = \{p_i, \mathbf{q}_i\}, i=1, L, M$.

To estimate the parameters from a given data set, we wish them in some sense best match the distribution of the data set. This can be done by using Expectation-Maximization (EM) algorithm [14]. The basic idea of EM algorithm is to maximize the likelihood function. Given a data set $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, assuming that the datum are independent of each other, the GMM likelihood can be written as,

$$L = \prod_{i=1}^N p(\mathbf{x}_i | \mathbf{Q}), \quad (3)$$

Maximize the likelihood function L is equal to minimize negative log-likelihood function,

$$\begin{aligned} E &= -\log L \\ &= -\log \prod_{i=1}^N p(\mathbf{x}_i | \mathbf{Q}) \\ &= -\sum_{i=1}^N \log p(\mathbf{x}_i | \mathbf{Q}) \\ &= -\sum_{i=1}^N \log \left(\sum_{k=1}^M p_k g(\mathbf{x}_i | \mathbf{q}_k) \right) \end{aligned} \quad (4)$$

The EM algorithm performs an iterative approximation to estimate the parameters which maximize the equation (4). The procedure is as follows:

Step 1: Initialize the parameters $\mathbf{q}_i = \{\mathbf{m}_i, \mathbf{S}_i\}$ and mixture weights p_i for each component by using clustering algorithm such as K-means [15] algorithm.

Step 2: Compute the probability that produced by the component m for given \mathbf{x}_i . This posterior probability is given by

$$p(m | \mathbf{x}_i, \mathbf{Q}) = \frac{p_m g(\mathbf{x}_i | \mathbf{q}_m)}{\sum_{k=1}^M p_k g(\mathbf{x}_i | \mathbf{q}_k)}, \quad (5)$$

Step3: Estimate the parameters for each component:

$$p_m = \frac{1}{N} \sum_{i=1}^N p(m | \mathbf{x}_i, \mathbf{Q}), \quad (6)$$

$$\mathbf{m}_m = \frac{\sum_{i=1}^N p(m | \mathbf{x}_i, \mathbf{Q}) \mathbf{x}_i}{\sum_{i=1}^N p(m | \mathbf{x}_i, \mathbf{Q})}, \quad (7)$$

$$\mathbf{S}_m = \frac{\sum_{i=1}^N p(m | \mathbf{x}_i, \mathbf{Q}) (\mathbf{x}_i - \mathbf{m}_m)(\mathbf{x}_i - \mathbf{m}_m)^T}{\sum_{i=1}^N p(m | \mathbf{x}_i, \mathbf{Q})} \quad (8)$$

Step4: Repeat step 2 to 4 until convergence.

3. GMM with Spatial Prior

3.1. Distance-Based Segmentation

GMM-based segmentation just estimates the color distribution and does not consider the spatial locations of objects. Nieuwenhuis *et al.* [2] believe that the locations of user markers are highly correlated. Thus, they proposed a strategy to estimate kernel densities in the joint space of color and location.

Let $I: \mathbb{W} \rightarrow \mathbb{R}^d$ denote an input image defined on the domain $\mathbb{W} \subset \mathbb{R}^2$, the image is divided by n pairwise disjoint regions W_i by labeling $u: \mathbb{W} \rightarrow \{1, L, n\}$. Each region can be expressed as $W_i = \{x | u(x) = i\}$. The distance based likelihood can be expressed by

$$p(x | u(x) = i) = \frac{1}{m_i} \sum_{j=1}^{m_i} k_{r_i}(x - x_{ij}) \quad (9)$$

Where m_i denote the number of user-labeled pixels x_{ij} in region W_i . k_{r_i} denote the adaptive kernel functions with spatial kernel width $r_i(x)$. In practice, Gaussian kernels are used. $r_i(x)$ is proportional to the distance from x to nearest sample point x_{ij} in region W_i . It can be written as

$$r_i(x) = a |x - x_{ij}| \quad (10)$$

where a is a distance factor.

3.2. GMM with Spatial Prior

From above discussion we defined a new likelihood function of GMM with distance information of user-labeled pixels by

$$L = \prod_{i=1}^n \prod_{j=1}^{N_i} p(I(x_{ij}) | \mathbf{Q}_i, u(x) = i) \prod_{k=1}^{m_i} \frac{1}{m_i} k_{r_i}(x_{ij} - x_{ik}) \quad (11)$$

where the subscript i denote the region number, the subscript j denote the j th element in region W_i , x_{ik} denote the user-labeled pixel in region W_i , m_i denote the number of user-labeled pixels in region W_i , N_i denote the number of pixels in region W_i , \mathbf{Q}_i denote the parameters of GMM in region W_i , n denote the image is divided by n regions.

To maximization the likelihood, the minimization of the negative log-likelihood function is considering,

$$E = -\log L = -\sum_{i=1}^n \sum_{j=1}^{N_i} \log(p(I(x_{ij}) | \mathbf{Q}_i, u(x_{ij}) = i)) + \sum_{i=1}^n \sum_{k=1}^{m_i} \log\left(\frac{1}{m_i} k_{r_i}(x_{ij} - x_{ik})\right) \quad (12)$$

Now, redefine the data term f_i by the equation

$$3 \tag{13}$$

Rewrite the equation (11) in a continuous form

$$E = \int_{i=1}^n \int_W f_i(x) dx, \tag{14}$$

The equation (14) meets $\bigcup_{i=1}^n W_i = W$, $W_i \cap W_j = \emptyset$, $i \neq j$.

3.3. Minimization via Convex Optimization

Suppose we have known the parameters, minimize energy E is equivalent to a segmentation problem. In order to get smooth boundaries, regularization terms are added to equation (14). It can be rewritten as

$$E = \int_{i=1}^n \int_W f_i(x) dx + \sum_{i=1}^n \int_W |u_i - 1|, \tag{15}$$

This optimization problem (15) is equivalent to continuous Potts model [16]. In order to solve the problem, a convex optimization technique build upon recent developments in primal-dual approaches [3, 17-19] can be used. Let the labeling function $u \in BV(\Omega, \{0,1\})^n$ represent the n regions W_i , and it is defined by

$$u_i(x) = \begin{cases} 1, & \text{if } x \in W_i \\ 0, & \text{otherwise} \end{cases} \quad i = 1, \dots, n, \tag{16}$$

Subject to $\int_{i=1}^n u_i = 1$ on W , and the total variation (TV) is defined by

$$TV(u_i) = \int_W Du_i = \int_W \sup_{\mathbf{x} \in \mathbb{R}^2} u_i(\mathbf{x}) d\mathbf{x}, \tag{17}$$

With

$$B = \{\mathbf{x} = (x_1^1, x_1^2) \in C_c^1(\mathbb{R}^2)^n \mid \|\mathbf{x}_i\| \leq 1, x_i^1, x_i^2, i = 1, \dots, n\}, \tag{18}$$

Where $\|\mathbf{x}_i\| = \sqrt{(x_i^1)^2 + (x_i^2)^2}$.

With these notations, the convex approximation of (15) can be written as

$$\min_{\{u_i\}_{i=1}^n} \int_{i=1}^n \int_W f_i(x) u_i(x) dx + \sum_{i=1}^n TV(u_i) \tag{19}$$

For two-phase case, since $u_1 = 1 - u_2$, a single labeling function u can be used to present u_1 and u_2 , in [3] $u_1 = (1 + u) / 2$ and $u_2 = (1 - u) / 2$. The constraint becomes $u^2 = 1$, (19) can be rewritten as (20) by dropping constant term and constant coefficient.

$$\min_{u^2=1} \max_{\mathbf{x} \in B} \left\{ E(u, \mathbf{x}) = \int_W (f_1 - f_2) u dx + \int_W u(\mathbf{x}) d\mathbf{x} \right\}, \tag{20}$$

The global optimal u is the binary value ± 1 . In this paper, suppose $u = -1$ is corresponding to f_1 and $u = +1$ is corresponding to f_2 . We also assume $u = -1$ is corresponding to foreground and $u = +1$ is corresponding to background. The Numerical implementation is proposed in [3] and the algorithm is

Step1: Initialize $\mathbf{x}^0 = 0$, b , t , $l > 0$

Step2: Compute

$$u^l = - \frac{\int_W \mathbf{x}^l + l(f_1 - f_2)}{\int_W \mathbf{x}^l + l(f_1 - f_2) + b}, \tag{21}$$

Step3: Update \mathbf{x}

$$x^{n+1} = \frac{x^n - t? u^n}{1 + t? u^n}, \quad (22)$$

Step4: Repeat step 2 to 3 until convergence.

Step5: Set

$$u = \begin{cases} 1, & \text{if } u^n \geq 0 \\ -1, & \text{if } u^n < 0 \end{cases}, \quad (23)$$

3.4. Estimation of Parameters

After each iteration, the solution of (20) is obtained, the parameters of foreground and background GMM can be estimate by

$$p_{im} | I(x_{ij}), \mathbf{Q}_i = \frac{p_{im} \mathcal{G}(I(x_{ij}) | \mathbf{q}_{im})}{\sum_{k=1}^M p_{ik} \mathcal{G}(I(x_{ij}) | \mathbf{q}_{ik})}, \quad (24)$$

Where $x_{ij} \in W_i$, $i \in \{1,2\}$ denote the foreground region and background region respectively. Thus we have $|W_i| = 2i - 3$. $\mathbf{Q}_i = \{p_{ik}, \mathbf{q}_{ik}\}, k=1, \dots, M$ denote the parameters of GMM in region W_i . The mixture weighting coefficients can be estimated by

$$p_{im} = \frac{1}{N_i} \sum_{j=1}^{N_i} p_{im} | I(x_{ij}), \mathbf{Q}_i, \quad (25)$$

Where N_i denote the number of pixels in region W_i . Mean vector \mathbf{m}_{im} and covariance matrix \mathbf{S}_{im} can be estimated by

$$\mathbf{m}_{im} = \frac{\sum_{j=1}^{N_i} p_{im} | I(x_{ij}), \mathbf{Q}_i I(x_{ij})}{\sum_{j=1}^{N_i} p_{im} | I(x_{ij}), \mathbf{Q}_i} \quad (26)$$

$$\mathbf{S}_{im} = \frac{\sum_{j=1}^{N_i} p_{im} | I(x_{ij}), \mathbf{Q}_i (I(x_{ij}) - \mathbf{m}_{im})(I(x_{ij}) - \mathbf{m}_{im})^T}{\sum_{j=1}^{N_i} p_{im} | I(x_{ij}), \mathbf{Q}_i} \quad (27)$$

3.5. Segmentation Algorithm

Since connectivity is an important cue for a good segmentation, this paper uses the largest connected region from foreground as a training data set to estimate foreground GMM parameters after each computation of u . The whole segmentation algorithm in this paper is the following procedure:

Algorithm:

1. Set user-labeled pixels as foreground W_1 and background W_2 . Initialize the parameters of Gaussian densities $\mathbf{q}_{ik} = \{\mathbf{m}_{ik}, \mathbf{S}_{ik}\}, k=1, \dots, M, i = \{1,2\}$ and mixture weighting coefficients $p_{ik}, k=1, \dots, M, i = \{1,2\}$ for foreground and background. Choose the parameters: $a, b, t, l > 0$. Compute $r_i(x)$ according to (10).

2. Compute f_1 and f_2 by (13).
3. Compute u by (21), (22), (23). If u convergence, then exit algorithm.
4. Find the connected regions C_1, C_2, \dots, C_r in foreground W_1 . Then, choose the largest

regions C_m and set $W_1 \setminus C_m$.

5. Update mixture weighting coefficients p_{ik} , $k=1,L,M$, $i=\{1,2\}$ and the parameters of Gaussian densities $q_{ik} = \{m_{ik}, S_{ik}\}$, $k=1,L,M$, $i=\{1,2\}$ by (24), (25), (26), (27).

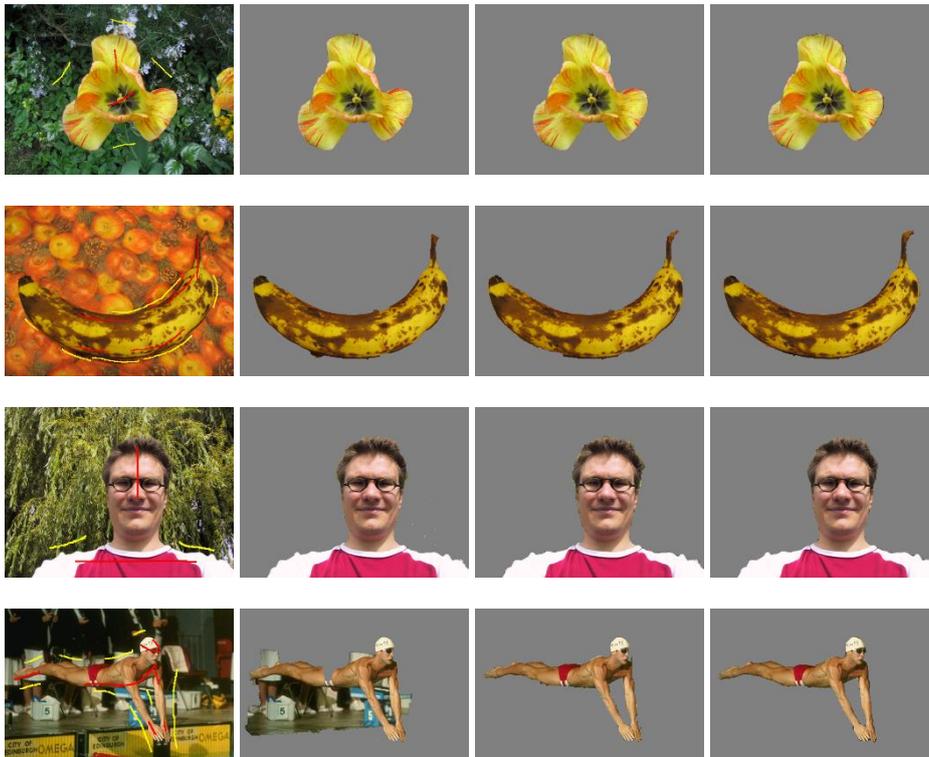
Repeat step 2 to 5.

4. Experimental Results

In this section, a large number of real-world images from public datasets such as GrabCut benchmark, Berkeley segmentation database, Graz benchmark [4] have been used to test the performance of the proposed method. In the experiments, we compare our method with well-known GrabCut method [1] and GSC (Geodesic Star Convexity) method [20]. The following parameters are used in this paper: $a = 5 : 20$, $b = 0.1$, $t = 0.01$, $l = 0.5 : 2$, $M = 5$.

4.1. Comparison Results on GrabCut Benchmark

In order to quantitatively compare the results, we use the most popular test images on GrabCut benchmark, then compute the true positive rate (TPR) and false positive rate (FPR) for these results which are shown in Figure 3. The TPR is defined as the ratio of the number of correctly classified pixels to the number of total pixels in the ground truth. The FPR is defined as the ratio of the number of mis-classified pixels in background to the number of total pixels in the ground truth. The higher the TPR is and the lower the FPR is, the better the method is.



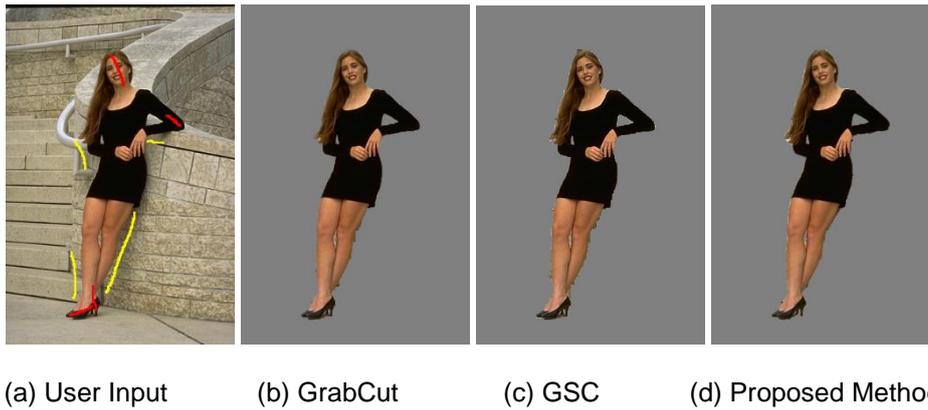


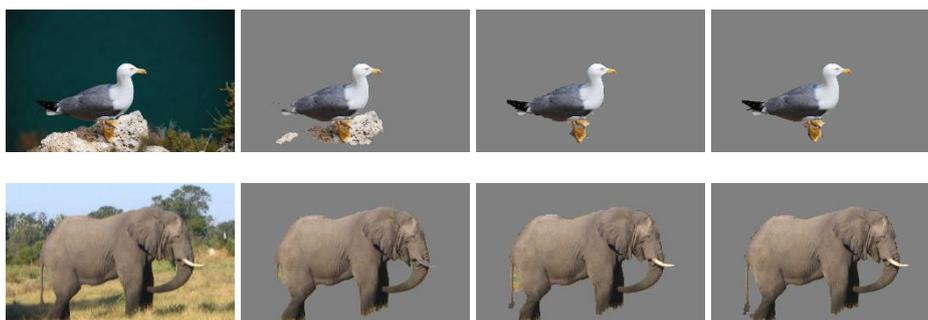
Figure 3. Comparison Results on GrabCut Benchmark between the GrabCut, GSC and Proposed Method. (a) User Input (b) Segmentation Result by GrabCut (c) Segmentation Results by GSC. (d) Segmentation Results by the Proposed Method

Table 1. The TPR and FPR Values of the Methods on GrabCut Benchmark

Input image	Method	TPR (%)	FPR (%)
flower	GrabCut	99.87	1.72
	GSC	99.82	3.02
	Proposed method	99.63	1.87
banana	GrabCut	98.16	1.82
	GSC	98.07	1.76
	Proposed method	98.81	1.80
person1	GrabCut	99.61	2.75
	GSC	99.61	3.57
	Proposed method	99.62	2.58
153093	GrabCut	92.33	157.59
	GSC	94.72	1.89
	Proposed method	95.66	1.43
388016	GrabCut	96.83	1.01
	GSC	97.61	1.98
	Proposed method	97.69	1.29

Table 1 lists the TPR and FPR results on five popular test images on GrabCut Benchmark. From these results, we can see that the proposed method has the higher TPR and the lower FPR in average.

In this Section, we use other test images from public datasets to test the performance of the proposed method in this paper. The segmentation results are shown in Figure 4.



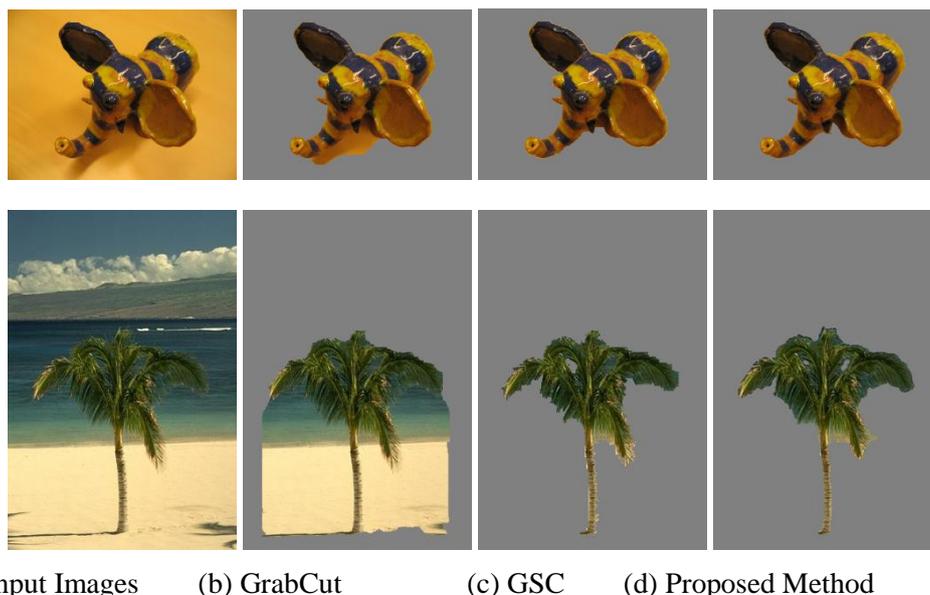


Figure 4. Comparison Results between the GrabCut, GSC and Proposed Method. (a) Input Images (b) Segmentation Result by GrabCut (c) Segmentation Results by GSC. (d) Segmentation Results by the Proposed Method

From Figure 4, we can see that since rough bounding boxes containing foreground and background inevitably, some foreground objects are surrounded by complex background can not be extracted accurately by GrabCut algorithm. GSC method introduce geodesic path and star-convexity prior in interactive image segmentation. So it has better segmentation performance than GrabCut. The proposed method in the paper combining the spatial locations of user markers with GMM improves the GrabCut's segmentation. Especially when the color distributions of foreground and background are overlapping, GrabCut may produce segmentation errors. Spatial locations will help to correct these errors. Since GSC method based on discrete Potts model and the proposed method based on continuous Potts model, we can see that the segmentation boundaries produced by the proposed method are smoother than GSC method.

5. Conclusion

This paper proposed an algorithm for interactive image segmentation. The method combines the spatial prior of user markers with GMM. Due to the spatial prior of user markers, the color distributions of foreground and background can be separated correctly when they are overlapping. The new likelihood function of GMM with distance information of user-labeled pixels is defined. The minimization of the negative log-likelihood function can be solved by a direct global minimization technique based on variational method. Since the method is based on continuous Potts model, the segmentation boundaries are smoother than those methods based on discrete Potts model such as GSC *etc.* on. Extensive experiments are performed to validate the proposed method in two-phase segmentation. The results show that the proposed method improves the segmentation performance of GrabCut. In the meantime, it gets the similar results as GSC and smoother segmentation boundaries. In the future, we will explore how to extend the method to multi-label segmentation and how to integrate texture feature into the method to get the better performance.

References

- [1] C. Rother, V. Kolmogorov and A. Blake, "ACM Trans. Graphics," vol. 23 no. 3, **(2004)**, pp. 309-314.
- [2] C. Nieuwenhuis and D. Cremers, *IEEE Trans. Patt. Anal and Mach Inte*, vol. 35 no. 5, **(2013)**, pp. 1234-1247.
- [3] Y. Gu, L. L. Wang, "IEEE Trans. Imag. Proc.," vol. 21 no. 5, **(2012)**, pp. 2399-2411.
- [4] J. Santner, T. Pock and H. Bischof, "Interactive Multi-Label Segmentation," *Proc. Asian Conf. Computer Vision*, **(2010)**.
- [5] D. Cermers, M. Rousson and R. Deriche, "Int. J. Comp. Visi.," vol. 72 no. 2, **(2007)**, pp. 195-215.
- [6] Y. Boykov and M. Jolly, "Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images," *Proc. Eighth IEEE Int. Conf. Computer Vision*, **(2001)**.
- [7] M. B. Salah and A. Mitiche, "IEEE Trans. Imag. Proc.," vol. 20 no. 2, **(2011)**, pp. 545-557.
- [8] T. F. Chan, S. Esedoglu and M. Nikolova, "SIAM J. Applied Math," vol. 66 no. 5, **(2006)**, pp. 1632-1648.
- [9] A. Chambolle, D. Cremers and T. Pock, "A Convex Approach for Computing Minimal Partitions," Technical Report TR-2008-05, Dept. of Computer Science, Univ. of Bonn, Germany, **(2008)**.
- [10] E. S. Brown, T. F. Chan and X. Bresson, "Int. J. Computer Vision, vol. 98 no. 1, **(2012)**, pp. 103-121.
- [11] P. Arbelaez, M. Maire, C. Fowlkes and J. Malik, "IEEE Trans. Pattern Analysis and Machine Intelligence," vol. 33 no. 5, **(2011)**, pp. 898-916.
- [12] J. F. Ning, L. Zhang, D. Zhang, C. K. Wu and *Patt. Reco.*, vol. 43 no. 2, **(2010)**, pp. 445-456.
- [13] C. Nieuwenhuis, E. Toppe and D. Cremers, *Int. J. Comp. Visi.*, vol. 104 no. 3, **(2013)**, pp. 223-240.
- [14] A. Dempster, N. Laird and D. Rubin, "J. of the Roya. Stat. Soci., Seri. B," vol. 39 no. 1, **(1977)**, pp. 1-38.
- [15] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proc. of the 5th Berkeley Symp. on Math. Stat. and Prob.*, vol. 1, **(1967)**, pp. 281-296.
- [16] R. B. Potts, "Some Generalized Order-Disorder Transformations," *Proc. Cambridge Philosophical Soc.*, vol. 48, **(1952)**, pp. 106-109.
- [17] A. Chambolle, "J. Math. Imag. Vis.," vol. 20 no. 1, **(2004)**, pp. 89-97.
- [18] J. Lellmann and C. Schnorr, "Continuous Multiclass Labeling Approaches and Algorithms," Technical Report, Univ. of Heidelberg, Germany, **(2010)**.
- [19] E. Bae, J. Yuan and X. C. Tai, "Int. J. Comp. Visi.," vol. 92 no. 1, **(2011)**, pp. 112-129.
- [20] V. Gulshan, C. Rother, A. Criminisi, A. Blake and A. Zisserman, "Geodesic Star Convexity for Interactive Image Segmentation," *CVPR*, **(2010)**.