Automatic GrabCut color Image Segmentation Based on EM Algorithm

LI Xiao-qi[1,a] , *LI Ye-li[2,b] and QI Ya-li[3,c]

[1,2,3]Department of Information Engineering,Beijing Institute of Graphic Communication,Xinghua Avenue(Band Two),Daxing,Beijing,China

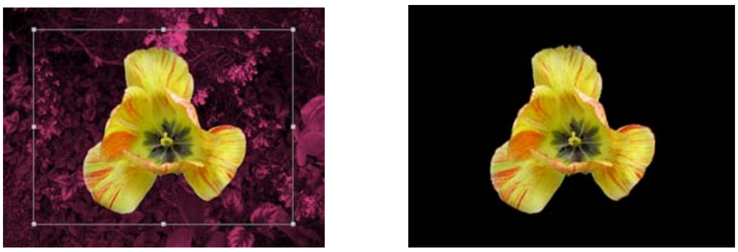
[a]lxq0102@gmail.com, [b]liyl@bigc.edu.cn, [c]qyl@bigc.edu.cn

Keywords: GrabCut, Expectation Maximization, Color Clustering, Gaussian Mixture model

Abstract. This paper presents a color image segmentation method using modified GrabCut. GrabCut is considered as one of semi-automatic image segmentation techniques, since it needs user interaction to initialize segmentation process. To eliminate this inconvenience, this paper propose using Expectation Maximization Algorithm for Gaussian mixture model to replace the original initialization process and achieve automatic GrabCut. After experiment and analysis, this paper gives an evaluation over speed and accuracy, proves practicability of this modification and provide a set up for further improvement.

1 Introduction

GrabCut [1] is an interactive color image segmentation based on Graph Cut [2] using adaptive GMMRF model in colors space[3]. GrabCut algorithm initialize the Gaussian mixture model(GMM)[4] with user interaction and developed an iterative optimization to replace the one-step algorithm in Graph Cut. Then the object boundary is processed with a robust "border matting" algorithm [5].The process of GrabCut color image segmentation from users' view is shown in Fig.1. Generally the user need just to select the object in the image with a rectangle, then the GrabCut will begin and give the result automatically.



(a) (b)

Figure 1 : GrabCut for color image segmentation process form users' view. (a) shows how the user interaction to select the object and (b) is the result of GrabCut segmentation.

As a state-of-the-art techniques for color image segmentation, a lot of research was done for GrabCut modification, improvement and application. Multi-scale Gaussian Smooth[6] and GMM optimization[7] are introduced to improve the robust and accuracy, at the cost of efficient of original GrabCut. Some other works use super-pixel method[8] or wavelet transform[9] to reduce data size. All the works mentioned above aim at speeding the algorithm [8,9] or improve its accuracy[6,7], but care less about the inconvenience and a probably poor segmentation[10].

Dina Khattab et al. proposed Automatic GrabCut based on clustering technique[10],along with Automatic GrabCut using different color space model was tested[11] and multi-label Automatic GrabCut for image segmentation[12]. According to their data, though some samples' performance are not well, the average error rate and time consume are both lower than the original GrabCut [10]. Inspired by the clustering techniques used to modify Grabcut, this paper present using Expectation Maximization algorithm to train GMM for color clustering and initialize GrabCut automatically.

2 Original GrabCut for color image segmentation

2.1 Graph Cut for monochrome image segmentation

Graph Cut described a segmentation as a pixel labeling problem[3]. Given image array $\mathbf{z}=(z_{1,\dots}z_n,\dots z_N)$, where n is the index of each pixel. Segmentation is to assign an "opacity" [3] value α_n to each pixel, denote as $\alpha=(\alpha_{1,\dots}\alpha_{n,\dots}\alpha_N)$. In foreground and background segmentation problem, $\alpha_n\in\{0,1\}$. Graph Cut describe the grey-level distributions of foreground and background with histograms respectively form labelled pixels:

$$\underline{\theta}=\{h(z;\alpha),\alpha=0,1\}. \tag{1}$$

The segmentation now become a problem to assign α_n to pixel n, whose value is z_n, with a given model $\underline{\theta}$. Then the algorithm define a "Gibbs" energy function to solve segmentation:

$$E(\underline{\alpha},\underline{\theta},z)=U(\underline{\alpha},\underline{\theta},z)+V(\underline{\alpha},z). \tag{2}$$

where term $\mathbf{U}$ is defined as:

$$U\left(\underline{\alpha},\underline{\theta},z\right)=\sum_n -\log h\left(z_n;\alpha_n\right). \tag{3}$$

Term U indicates the fit of distribution $\underline{\alpha}$ to the image date $\mathbf{z}$ in the form of the sum of negative logarithm of probability function, given model $\underline{\theta}$. So when the fit is the best, term $\mathbf{U}$ is minimized.

Term V is defined as:

$$V\left(\underline{\alpha},z\right)=\gamma\sum_{(m,n)\in C}\mathrm{dis}(m,n)^{-1}\delta\left(\alpha_n,\alpha_m\right)\exp(-\beta(z_m-z_n)^2)\,.\,\delta\left(\alpha_n,\alpha_m\right)=\begin{cases}1 & \text{if } \alpha_n\neq\alpha_m\\ 0 & \text{otherwise}\end{cases}. \tag{4}$$

where γ and β are constant chosen according to experimental data [1,2],C is the set of pairs of neighboring pixels and function dis(m,n) is the Euclidean distance of indexed pixels. This term is smoothness term and indicates the penalty for discontinuity between pixel n and pixel m[1]. Hence the term V is negative correlated with $|z_{m}-z_n|$.

So given model $\underline{\theta}$, when each values z_n of pixel n are best fit and the pairs of neighboring pixels n and m with different α_n and α_m assigned are very different in their values z_n and z_m, the function reaches its global minimum. Hence the best segmentation can be obtained by estimating the global minimum of energy function. A minimum cut algorithm [1,13] is applied for minimization.

2.2 GrabCut algorithm

In GrabCut the model for monochrome images is replaced by Gaussian mixture model (GMM) for colors images. According to practise of soft segmentation[14,5],a new vector $\mathbf{k}=\{k_1,\dots,k_n,\dots,k_N\}$ is introduced and assigned to each pixel belonging to GMM's k_nth component, $k_n=1,2,\dots,K$ (Normally K = 5)[2], and $\alpha_n = 0,1$ is assigned to each pixels to indicate that it belongs to either foreground GMM or background GMM. So the energy function now become:

$$E(\underline{\alpha},k,\underline{\theta},z)=U(\underline{\alpha},k,\underline{\theta},z)+V(\underline{\alpha},z). \tag{5}$$

and GMM is defined as[4]:

$$G(z)=\sum_{i=1}^{K}\omega_k g_k(\mathbf{z};\mathbf{u}_\mathbf{k},\Sigma_\mathbf{k}),\sum_{i=1}^{K}\omega_k=1,\text{and}\,0<\omega_k<1. \tag{6}$$

Where $g_k=(\mathbf{z};\boldsymbol{\mu}_\mathbf{k},\boldsymbol{\Sigma}_k)$ is the Gaussian distribution function for each component k, k=1,2,…K:

$$g\left(\mathbf{z};\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=\frac{1}{\sqrt{2\pi^D|\boldsymbol{\Sigma}|}}\exp\left[-\frac{1}{2}\left(\mathbf{z}-\boldsymbol{\mu}\right)^T\Sigma^{-1}\left(\mathbf{z}-\boldsymbol{\mu}\right)\right]. \tag{7}$$

and ω_k is the weighting coefficient, vector $\boldsymbol{\mu}_\mathbf{k}$ is the means and $\boldsymbol{\Sigma}_k$ is the covariance matrix for kth component and D is the number of dimension of variable **z**.

Combine Eq. 5 and Eq. 6, term U now become:

$$U\left(\underline{\alpha},k,\underline{\theta},z\right)=\sum_{n}G\left(\underline{\alpha},k,\underline{\theta},z\right). \tag{8}$$

Where

$$G\left(\underline{\alpha},k,\underline{\theta},z\right)=-\log\omega\left(\alpha_n,k_n\right)+\frac{1}{2}\log\left|\Sigma\left(\alpha_n,k_n\right)\right|+\frac{1}{2}\left[z_n-\mu\left(\alpha_n,k_n\right)\right]^T\Sigma\left(\alpha_n,k_n\right)^{-1}\left[z_n-\mu\left(\alpha_n,k_n\right)\right]. \tag{9}$$

And model $\underline{\theta}$ now become:

$$\underline{\theta}=\left\{\pi\left(\alpha,k\right),\mu\left(\alpha,k\right),\Sigma\left(\alpha,k\right),\alpha=0,1,k=1,\ldots,K\right\}. \tag{10}$$

As for term V, the Euclidean distance now is a 2-norms of z_{m}- z_n:

$$V\left(\underline{\alpha},z\right)=\gamma\sum_{(m,n)\in C}\delta\left(\alpha_n,\alpha_m\right)\exp\left(-\beta\left\|z_m-z_n\right\|^2\right). \tag{11}$$

GrabCut minimize the energy function with a modified iterative minimization cut algorithm instead of the one-step minimum cut. The algorithm initialized with user interaction to two set of pixel, one for background where α_n=0 and one as possible candidates of object where α_n=1. Two GMM were initialized respectively with the two sets. Then, the iteration begins:

1. The algorithm first assign k_n to each pixel n: $k_n = \arg\min_{k_n} G(\alpha_n, k_n, \theta_n, z_n)$;
2. Then update the GMM parameters from data **z**: $\underline{\theta} = \arg\min_{\underline{\theta}} G(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z})$;
3. At last, Estimate minimization using ordinary min cut and reassign $\underline{\alpha}$ and **k**;

Repeat these steps until the result convergence and the segmentation is done.

3 Automatic GrabCut based on EM algorithm

3.1 EM algorithm for automatic GrabCut

One of the disadvantages of GrabCut is that the segmentation is initialized by user interaction which may leads to bad segmentation if initialization quality is poor[11]. Rother[2] proposed further user editing to improve the segmentation but require more user interaction. To achieve a fully automatic algorithm, a method using EM algorithm to train GMM is proposed.

Expectation Maximization algorithm [15] is a maximum likelihood estimate algorithm, using a hidden random variable and Jensen inequality to iteratively estimate the parameters that maximize the probability function. In this case, the parameters to be estimated to clustering pixels into 2 sets are two new GMM parameters [16], one for foreground and one for back ground. Thus require $k \geq 2$ for components. Since it require 2 sets to initialize the GrabCut and when $k > 2$ it is hard to tell components belong to foreground or background[17],so k=2 is fixed.

3.2 iterative estimation in EM algorithm

With Eq. 10, the iterative estimation for GMM parameters is as followed:

Step 1. Initialize GMM $\underline{\theta}^{(t)}$as $\underline{\theta}^{(0)}, \underline{\theta}^{(0)} = \{\boldsymbol{\mu}(k)^{(0)}, \boldsymbol{\Sigma}(k)^{(0)}, \boldsymbol{\omega}(k)^{(0)}\}$,where k = 1,2;

Step 2. Expectation step (E-step): Compute $\varphi(\theta;\underline{\theta}^{(t)}) = \sum_{x} P(x \mid z, \underline{\theta}^{(t)}) \log P(z, x \mid \underline{\theta}^{(t)})$, where x is the hidden variable.

Step 3.Maximization step (M-step): Compute $\underline{\theta}^{(t+1)} = \arg\max_{\underline{\theta}} \varphi(\theta; \underline{\theta}^{(t+1)})$, and update parameters.

Repeat step 2 and step 3 until convergence, then go to next step.

Step 4. Assign α_n to each pixel n according to the best estimation of $\underline{\theta}$, clustering image into two sets.

Then GrabCut can be initialized with these two sets, one for foreground and one for background, instead of manually initialization and GrabCut will automatically begins.

4 Conclusion and discussion

Both original GrabCut and proposed method are tested and the results of segmentation are compared. Some of the typical tests are shown in Fig 2.

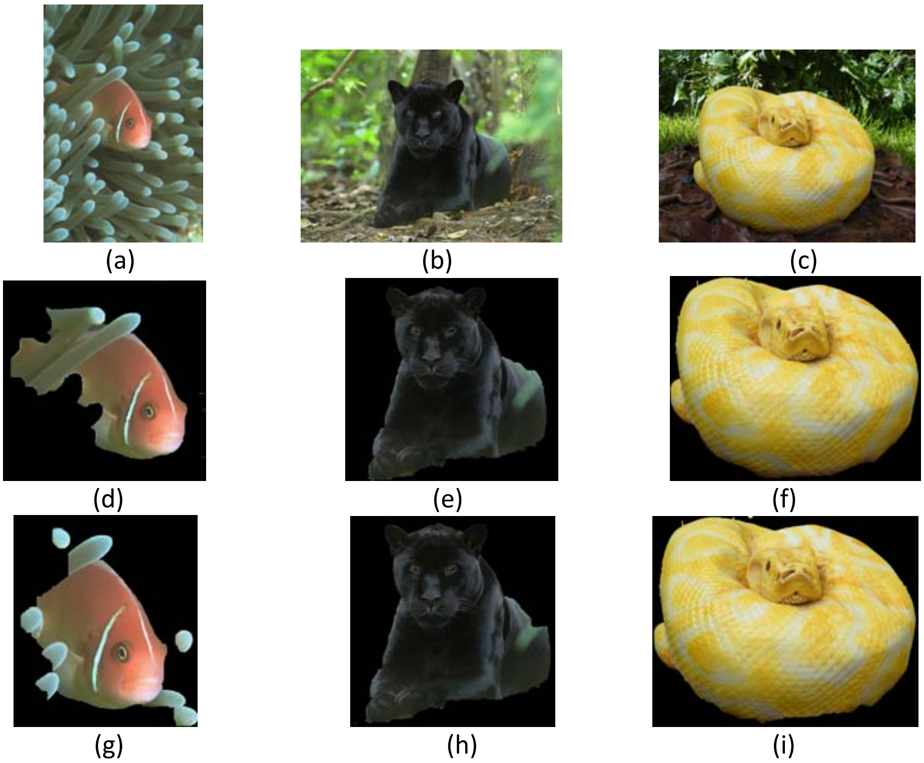


Figure 2 (a),(b),(c)is the original image (d),(e),(f)is the segmentation using original Grabcut(without border mating and further user editing) (g),(h),(i)is the segmentation using proposed method in this paper.

And the time consume and accuracy are shown respectively in Table 1 and Table 2.

Table 1 Time consume

	Average-Total/ms	Avg-1st iteration /ms	Avg-2nd iteration /ms	Avg-3rd iteration /ms	Avg-4th iteration /ms	Avg-5th iteration /ms
Oringinal GrabCut	24767.56	5337.26	3463.34	3315.25	3201.79	3147.65
Modifed GrabCut	25602.34	4853.67	3797.23	3504.76	3314.47	3154.24

Table 2 Accuracy

	Original GrabCut	Modified GrabCut
Accuracy%	92.32%	90.51%

With a slight decrease in accuracy and a little increase in time consume, the proposed modification achieved an automatic GrabCut. However, Fig 2 (d), (g) shows segmentation between two methods can be far more different. This is occurred by the different initialization form user interaction and EM algorithm. User interaction takes the advantage of human's detect ability that could always select the object pixels in the image. EM algorithm somehow would convergence into local optimum [18] when dealing with the global maximization and select background pixels that is similar to the foreground pixels into foreground model. That will cause large data size and accuracy decrease, resulting in what is shown in two table. Since the GrabCut is sensitive to initialization [11], more works for a better clustering needs to be done in further research.

Acknowledgments

This work was supported in part by the special subject of Beijing Institute of Graphic Communication, Program funded by the Education Commission of Beijing (TJSHG201310015016).

References

[1]Rother C, V. Kolmogorov, and A. Blake, ""GrabCut": interactive foreground extraction using iterated graph cuts," ACM Transactions on Graphics, vol. 23, no. 3, pp. 309–314, 2004.
[2] BOYKOV, Y., AND JOLLY, M.-P. 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In Proc. IEEE Int. Conf. on Computer Vision,
[3]BLAKE, A., ROTHER, C., BROWN, M., PEREZ, P., AND TORR,P. 2004. Interactive Image Segmentation using an adaptive GMMRF model. In Proc. European Conf. Computer Vision.
[4]Chris Stauffer, W.E.L Grimsos. Adaptive Background mixture models for real-time tracking. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Comput. Soc. Part Vol 2.1999.
[5] CHUANG, Y.-Y., CURLESS, B., SALESIN, D., AND SZELISKI, R.2001. A Bayesian approach to digital matting. In Proc. IEEE Conf. Computer Vision and Pattern Recog., CD–ROM.
[6] Tomoyuki Nagahashi, Hironobu Fujiyoshi, Takeo Kanade Image Segmentation Using Iterated Graph Cuts Based on Multi-scale Smoothing, ACCV 2007, Part II, LNCS 4844, pp. 806–816, 2007
[7]Chen D, Chen B, Mamic G, et al. Improved GrabCut segmentation via GMM optimization. Proc. of the 2008 International Conference on Digital Image Computing: Techniques and Applications. Washington, DC: IEEE Computer Society. 2008. 39-45.
[8] HU Zhi-li, GUO Min. Fast segmentation in color image based on SLIC and GrabCut[J/OL].Computer Engineering and Application
[9] LIU Huan-Huan, Yao Ming-Hai, Wang Xian-Bao. GrabCut Image Segmentation Based on Wavelet Transform [J] Computer System Application.2014,23(8):154-157
[10]D. Khattab, H. M. Ebied, A. S. Hussien, andM. F. Tolba, "Automatic GrabCut based on unsupervised clustering for image segmentation," Intelligent Systems'2014 Advances in Intelligent Systems and Computing Volume 323, 2015, pp 579-592
[11] Khattab D. , Ebied H.M.; Hussein A.S. , Tolba M.F. ,Color Image Segmentation Based on Different Color Space Models Using Automatic GrabCut. The Scientific World Journal,Volume 2014, Article ID 126025,10 pages
[12] Khattab, D.; Ebied, H.M.; Hussein, A.S.; Tolba, M.F., "Multi-label automatic GrabCut for image segmentation," *Hybrid Intelligent Systems (HIS), 2014 14th International Conference on* , vol., no., pp.152,157, 14-16 Dec. 2014
[13] Boykov, Y., and Kolmogorov, V. 2003. Computing Geodesics and Minimal Surfaces via Graph Cut. In Proc. IEEE Int. Conf. on Computer Vision.
[14]RUZON M and TOMASI C. Alpha estimation in natural images. IEEE Conf.Comp.Vision and Pattern Recog.2000
[15] A.P. Dempster, N.M. Laird, and D.B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM algorithm", *Journal of the Royal Statistical Society*, Series B, vol. 39, 1:1-38
[16] Redner RA, Walker HF. Mixture density, maximum likelihood and the EM algorithm [J]. SIAM Review , 1984, 26(2): 195-239
[17] FENG C. Research of K-means Clustering Algorithm [D].DALIAN University of Technology.2007.
[18]ZHONG Jinqin, GU Lichuan, TAN Jieqing, LI Yingying. Estimating parameters of GMM based

on split EM. Computer Engineering and Applications, 2012,48(34):28-32.