Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp50-55)

MRF Matting on Complex Images

Shengyou Lin[1], **Ruifang Pan**[1], **Jiaoying Shi**[2]
[1]*Dept. of Communicative Technology, ZJICM, Hangzhou, P.R.C.*

[2]*State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, P.R.C.*

Abstract

This paper proposes a novel approach to solve the matting problem on complex images using Markov Random Field(MRF) model. Although many natural image matting methods have been proposed, matting on complex images still remains as a challenge. Our approach, which we call MRF matting, partitions the image manually into three regions: foreground, background, and unknown region. Then, the unknown region is roughly segmented into several joint sub-regions by the user. In each sub-region, matting labels are defined and modelled as an MRF and assigned to the pixels in unknown region. Matting problem is then formulated as a maximum a posteriori(MAP) estimation problem on this MRF model and its associated Gibbs distribution. Simulated annealing is used to find the optimal matting solution. We compute alpha mattes of all the sub-regions and combine them into a final matte. When matting on complex images, our approach is demonstrated to be more robust than existing methods. Experimental results are shown and compared with other methods in this paper.

Key Words: *Markov Random Field, natural image matting, image segmentation, maximum a posteriori, simulated annealing*

1. Introduction

The matting problem is to separate of a fractional alpha value $\alpha(0\leq\alpha\leq1)$, a foreground color component F, and a background color component B from the color C of any pixel in a given image. If α is binary valued ($\alpha\in\{0,1\}$), this kind of matting problem is also called image segmentation. Matting is a hot research topic in recent years and used widely in the film and video production to make special effects.

The matting techniques can be classified mainly as *blue screen matting* and *natural image matting*. If the background color component B of every pixel is constant, this matting problem is called blue screen matting, while if B is arbitrary, it will be natural image matting. Matting is very difficult because it is essentially an under-constrained problem. Therefore, additional constraints or user-interactions are required. Blue screen matting has been summarized nicely by Smith and Blinn in their paper[1]. Natural image matting attempts to pull a matte from an arbitrary background using three basic steps: segmenting the image into three regions, namely, foreground, background, and unknown; and estimating the background and foreground color components of each pixel in unknown regions; estimating the alpha value of this pixel. Most of natural image matting methods [2,3,4,5,6,7] work well on smooth image. Today researchers begin to pay attention to pulling a matte from complex scene, such as local Poisson matting[6]. In this paper, we propose a new MRF model-based matting approach to pull the matte from complex image.

The remainder of the paper is organized as follows. We give a review of the existing natural image matting in Section 2. Our MRF matting method is introduced in Section 3. We show the results of some examples in Section 4 and give some discussions about our approach in Section 5. Our work is summarized and some future research directions are point out in Section 6.

2. Previous Work

Blue screen matting has been used in the film and video industry for decades. Main limitation of blue screen matting is the reliance on a controlled background image. Natural image matting is generally composed of three steps: region segmenting, color estimating, and alpha estimating. Most of the natural image matting methods [2,3,4,5,6,7] begin from a user-supplied trimap which segments the image into three regions: definitely foreground, definitely background and unknown region.

In Knockout[2], the estimated F and B are weighted mean of the colors of pixels along the perimeter of the foreground and background regions. The final estimated α is a weighted mean of three alpha components calculated in R,G,B channels. Knockout method is used widely in industry since it is fast. In [3] and [4], the color samples for F and B are analyzed by PCA and a mixture unoriented Gaussians respectively. Bayesian matting[5] partitions the foreground and background color samples into some clusters. Each cluster is fitted with an oriented Gaussian distribution. An MAP estimation of α, F and B is calculated simultaneously in a Bayesian framework. Once F and B are computed, these methods use a same alpha estimating method to compute α by projecting color C onto the line segment FB in RGB color space. Bayesian matting is an important development for natural image matting. It can obtain better results than previous methods. One weakness of these statistical matting algorithms is that they are slow.

Poisson matting[6] is the most important progress for natural image matting after Bayesian matting. Global Poisson matting works well in smooth images and obtains comparable results with Bayesian matting in many examples. More importantly, if the results of global Poisson matting are not satisfactory, local Poisson matting can be introduced to refine them using filtering tools. Instead of editing a trimap like most natural image matting approaches, Poisson matting provides a completely different way to adjust a matte by modifying image gradient field. Unlike other methods to estimate alpha in RGB color space, we presented an efficient matting method[7] in perceptual color space, called perceptual matting here. This method separates the chroma and intensity information of a color and emphasizes the more significant one. When the image is smooth, perceptual matting produces modestly better matte than Bayesian matting and can extract the foreground object as fast as Knockout method without obvious weakening the matting quality.

However, all these methods [2,3,4,5,6,7] are not adequately robust for complex images. The main limitation of [2,7] is that they use a weighted-mean color estimating method. When the image is complex, F and B obtained by this color estimating method will largely bias the true values and can't be used in the following alpha estimating. The common problem of [2,3,4,5] is that they adopt an inadequate alpha estimating method. Their alpha estimating methods will introduce problems in some cases, which is explained detailedly in [7]. Global Poisson matting is also not robust for complex images and will result in a poor quality mattes in this case. Although local Poisson matting can refine these mattes using additional tools, some limitations still remain in Poisson matting. First, operations in Poisson matting are based on a gray image that is converted from the original color image.

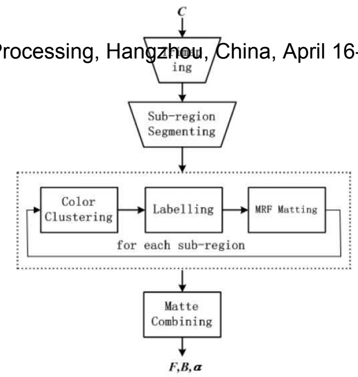


Figure 1: MRF matting flow chart. The input image C is segmented manually into three regions: foreground, background and unknown regions. Then the unknown region is partitioned into several smaller joint sub-regions. In each sub-region, the foreground and background pixels are clustered. After labelling and MRF matting, F, B and α of every pixel in this sub-region are estimated simultaneously. Finally, all the mattes of sub-regions are then finally combined into a final matte.

This conversion inevitably loses some information of the original image. Second, although local Poisson matting refines the matte, it increases extra user interactions at the same time.

Here, a complex image means an image with high resolution or high color-variation in a small region. There are a large numbers of complex images around us. It is necessary to find a matting approach to extract objects from these images. We call this problem complex image matting. How to remove the limitations of previous methods while still preserving their advantages in a largest scale will be always a problem in complex image matting.

In this paper, we introduce a novel complex image matting method called MRF matting. This method produces better results than global poisson matting and perceptual matting and requires less user interactions than local poisson matting.

3. MRF Matting

Like previous matting techniques, our approach starts from manually segmenting an input image into three regions: foreground, background, and unknown region. Then, the unknown region is segmented into several joint sub-regions by the user. In each sub-region, matting labels are defined and modelled as an MRF and assigned to pixels in unknown region. Matting problem is then formulated as an MAP estimation problem on this MRF model and its associated Gibbs distribution. We compute alpha mattes of all the sub-regions and combine them into a final matte. Furthermore, we use a special matting approach to extract foreground ob-

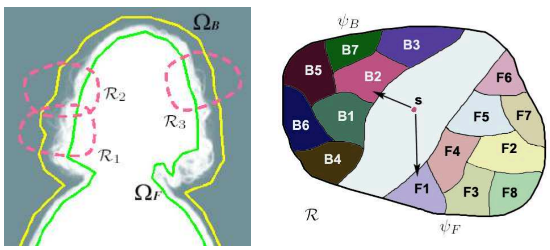
Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp50-55)

jects, which obtains better results than MRF matting. Figure 1 illustrates the flowchart of MRF matting.

3.1. Matting for General Images

Recently, MRF model has attracted much attention in computer vision community. A detailed description on MRF modelling can be found in [8] or [9].

In MRF matting, we first segment a complex image into definitely foreground, definitely background and unknown region using hand-drawn foreground contour Ω_F and background contour Ω_B. The unknown region is then segmented into several sub-regions. Let $\mathcal{R}$ denote a sub-region, ψ_F and ψ_B denote the color set of the foreground and background pixels in $\mathcal{R}$ respectively. We partition ψ_F, ψ_B into M and N clusters using the method of Orchard and Bouman[10], as shown in Figure 2. Let $F_i(i\in\{1,...,M\})$ denote the weighted mean of colors in the ith foreground cluster, $B_j(j\in\{1,...,N\})$ the weighted mean of colors in the jth background cluster.



(a) (b)

Figure 2: Regions segmenting and color estimating. (a) shows foreground contour Ω_F , background contour Ω_B, and three subregions: $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_3$. And (b) shows sub-region $\mathcal{R}$ in detail. In $\mathcal{R}$, the color set ψ_F of foreground pixels is partitioned into 8 clusters and the color set ψ_B of background pixels is partitioned into 7 clusters. Given a pixel s in $\mathcal{R}$, MRF matting system finds its foreground and background color components F_1 and B_2. The partitions F_1-F_8 or B_1-B_7 are actually in color space, not in the x-y space. In fact, clusters F_1 etc can represent pixels which are maybe spatially far away.

Matting problem can be posed as a labelling problem. Assume that there are K unknown pixels in $\mathcal{R}$. Let $\Lambda=\{1,...,K\}$ index a set of K sites. Each site s is one unknown pixel in $\mathcal{R}$ and $\mathcal{G}_s(\mathcal{G}_s\subset\Lambda)$ is the 4 neighborhood system of s. For each pair of F_i and B_j, the corresponding alpha value $\alpha^s_{(i,j)}$ can be calculated using perceptual alpha estimation method. In MRF matting, a label is a triple $(F_i,B_j,\alpha^s_{(i,j)})$. The labelling problem is to assign a label x_s to each site s. A solution $X=\{x_s|s\in\Lambda\}$ is called a configuration.

In Bayesian matting, the likelihood L_α is assumed as a constant and its definition derived from statistics of real alpha mattes is left as future work. This neglect of L_α will lead to a poor matte with jagged edges, as pointed out in [11]. In the following, we'll redefine the likelihoods using MRF model.

Let $\mathbf{C}$ denote the original image in $\mathcal{R}$. The foreground image $\mathbf{F}$ and background image $\mathbf{B}$ of $\mathbf{C}$ are independent to each other. The matte $\mathbf{A}$ of $\mathbf{C}$ is completely dependent on $\mathbf{F}$ and $\mathbf{B}$ because $\mathbf{A}$ can be derived from $\mathbf{F}$ and $\mathbf{B}$ using perceptual alpha estimating method. Therefore, we only need to find the most likely estimates for $\mathbf{F}$ and $\mathbf{B}$. This can be expressed as follows using Bayes's rule:

$$\begin{aligned} &\arg\max_{\mathbf{F},\mathbf{B}} P(\mathbf{F},\mathbf{B}|\mathbf{C}) \qquad (1)\\ =&\arg\max_{\mathbf{F},\mathbf{B}} P(\mathbf{C}|\mathbf{F},\mathbf{B})P(\mathbf{F})P(\mathbf{B}|\mathbf{F})/P(\mathbf{C})\\ =&\arg\max_{\mathbf{F},\mathbf{B}} P(\mathbf{C}|\mathbf{F},\mathbf{B})P(\mathbf{F})P(\mathbf{B}). \end{aligned}$$

At the same time, $P(\mathbf{F},\mathbf{B}|\mathbf{C})$ can be written as a Gibbs distribution, dropping the constant term $P(\mathbf{C})$:

$$P(\mathbf{F},\mathbf{B}|\mathbf{C}) \propto \frac{1}{Z}\exp -E,$$

where

$$E = U(\mathbf{C}|\mathbf{F},\mathbf{B}) + U(\mathbf{F}) + U(\mathbf{B}). \qquad (2)$$

The solution of MRF matting can be obtained by minimizing a Gibbs energy E. The key step is to define the likelihood energy $U(\mathbf{C}|\mathbf{F},\mathbf{B})$ and the prior energy $U(\mathbf{F})$ and $U(\mathbf{B})$ of a configuration X.

Likelihood energy. Because we use a perceptual approach to estimate alpha, consequently, we must find a perceptual strategy to evaluate the cost of a matting label. In perceptual alpha estimating method, the chroma and intensity information of a RGB color are separated in perceptual color space and the more significant one is emphasized. Take site s in $\mathcal{R}$ for example, its intensity alpha α^s_{IN} and chroma alpha α^s_{CH} are estimated respectively. After their weight W^s_{IN}, W^s_{CH} are computed, the weighted mean of these two alpha values yields the final alpha value α_s. Hence, we can define a ratio ε_{x_s} as the cost of label $x_s=(F_i,B_j,\alpha^s_{(i,j)})$, for pixel s in X as follows:

$$\varepsilon_{x_s} = \begin{cases} W^s_{IN}/W^s_{CH} & \text{if } W^s_{IN}<W^s_{CH} \\ W^s_{CH}/W^s_{IN} & \text{if } W^s_{IN}\geq W^s_{CH}. \end{cases}$$

The smaller ε_{x_s} is, the more convincing this estimated alpha is. So, the likelihood energy of a configuration can be given as a sum of the cost of label :

$$U(\mathbf{C}|\mathbf{F},\mathbf{B}) = \sum_{s\in\Lambda} \varepsilon_{x_s}. \qquad (3)$$

Prior energy. Let $F_s(s\in\Lambda)$ denote the foreground color of s in X. We simply use the Euclid distance of

RGB color space to measure the prior energy of $\mathbf{F}$,$\mathbf{B}$:

Proceedings of the 6th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 16-18, 2006 (pp50-55)

$$U(\mathbf{F}) = \lambda_F \sum_{s\in\Lambda} \sum_{r\in\mathcal{G}_s} \|F_s - F_r\| \quad (4)$$

$$U(\mathbf{B}) = \lambda_B \sum_{s\in\Lambda} \sum_{r\in\mathcal{G}_s} \|B_s - B_r\| \quad (5)$$

where λ_F, λ_B are adjustable influence parameters used to control the influence of $U(\mathbf{F})$ and $U(\mathbf{B})$. In our implementation, λ_F and λ_B are both initially set to 1 and can be adjusted to other values according to different images and sub-regions.

Line process. When the foreground object has a hard edge, MRF process will introduce "Manhattan" artifacts in which the border of the foreground object in matte often fail to correspond to the edge in original image. To reduce the effect of "Manhattan" artifacts, a line process $\mathbf{L}$[12], can be adjoined to MRF model. Let d denote a line site between site r and site s. if l_d=1, an edge element is "present" at d and the bond between r and s is "broken". The likelihood energy is given the same as before but the prior energy is changed to be composed of three terms, say $U(\mathbf{F}|\mathbf{L})+U(\mathbf{B}|\mathbf{L})+U(\mathbf{L})$. These three terms can be modelled as follows:

$$U(\mathbf{F}|\mathbf{L}) = \lambda_F \sum_{s\in\Lambda} \sum_{r\in\mathcal{G}_s} \|F_s - F_r\|(1 - l_{s,r}) \quad (6)$$

$$U(\mathbf{B}|\mathbf{L}) = \lambda_B \sum_{s\in\Lambda} \sum_{r\in\mathcal{G}_s} \|B_s - B_r\|(1 - l_{s,r}) \quad (7)$$

$$U(\mathbf{L}) = \lambda_L \sum_{c\in\mathcal{C}} V_c \quad (8)$$

where the influence parameter λ_L is set to be varied with the change of λ_F and λ_B:

$$\lambda_L = 255.0 \times (\lambda_F + \lambda_B)/2.0.$$

Energy minimizing. There are totally $(MN)^K$ configurations to the matting problem in $\mathcal{R}$. For a moderate value of K=600, supposed that there are only 4 possible alpha values for every unknown pixel, there will be up to 4^{600} solutions! The identification of even near-optimal solution will be extremely difficult. It is definitely necessary to partition the unknown region into some smaller sub-regions and find the globally optimal solution in each sub-region. There is a relatively efficient algorithm–simulated annealing[13]–for finding the globally optimal realization in MRF application. Our implementation of simulated annealing in MRF matting lowers the temperature by a factor $\tau = 0.95$.

After the mattes of all sub-regions are computed, they are merged to a final matte. For pixels in the overlapped area between two neighboring sub-regions, we simply take the average of alpha values in two sub-regions as the final alpha values of these pixels.

4. Results

As pointed out in Section 2, perceptual matting can obtain better mattes than previous methods and can extract the foreground object as fast as Knockout method without obvious weakening the matting quality.. Here we compare our results with perceptual matting and global poisson matting with some complex images.

In Figure 3, MRF matting produces comparable results with perceptual matting and global Poisson matting when the image is smooth. The unknown region is segmented into 11 sub-regions in this example.

In Figure 4, given a trimap of complex image, the results of perceptual matting show more noises than those of MRF matting. There are several blocks of mismatting regions in the results of global Poisson matting, because the image gradient fields of these regions bias the matte gradient fields largely. The unknown regions of these five examples are partitioned respectively into 23, 15, 16, 25, and 21 sub-regions.

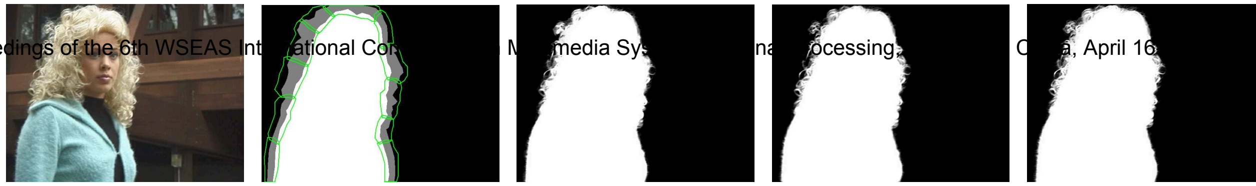
To reduce computation time, we partition the unknown region into smaller sub-regions. The number of sub-regions depends on the size and complexity of an image. Usually, the computation time in a sub-region varies from several seconds to more than 1 minute. For a 335×265 image with 15 sub-regions, as shown in the second example in Figure 4, it takes the computer(CPU: Pentimn IV 1.8G, RAM: 512M) more than 15 minutes to extract the foreground objects, including the time for user to specify trimap and sub-regions. A lot of works are left in energy minimizing to improve the efficiency of our method.

5. Discussions

Complex image matting is a very difficult and essentially underconstraint problem. Some problems still remain in MRF matting.

First, in alpha estimating, the first and so the most important step in MRF matting, perceptual alpha estimating method we used is not accurate for all cases. Although progresses have been made, perceptual alpha estimating method still needs improvements. For example, the distance of image space could be added as an influence factor into the analysis of intensity weight and chroma weight. Besides, the computing of likelihood and prior energy is another important step. An effective energy computing process enables the matting system to pick up the best solution from numerous configurations.

Second, MRF matting is still not robust when an image has complex foreground objects with hair-like silhouettes and complex background. Acceptable results can be obtained when the foreground or/and the



(a) Original (b) TRIMAP (c) MRF (d) Global Poisson (e) Perceptual

Figure 3: Comparison of matting results of smooth image. We segment the unknown region into 11 sub-regions. Under the same trimap in (b), the result of MRF matting(c) shows comparable quality with those of global Poisson matting(d) and perceptual matting(e).

(a) Original (b) TRIMAP&Sub-regions (c) MRF (d) Global Poisson (e) Perceptual

Figure 4: Comparison of matting results of complex images. The results of MRF matting exhibit less visible artifacts than those of global Poisson matting and perceptual matting. The unknown regions of these five examples are partitioned respectively into 23, 15, 16, 25, and 21 sub-regions. These green polygons show what the user-specified division into sub-regions looks like. (This Gandalf image is composited with a new background image and the wizard Gandalf, who is extracted from an image originally from movie "The lord of the Ring" using perceptual matting. Other images are obtained from http://www.research.microsoft.com/vision/cambridge/segmentation/ and http://www.cs.berkeley.edu/projects/vision/grouping/segbench/.)

background is/are smooth, such as the first example in Figure 4, where the foreground object is smooth and the background is high-resolution. MRF matting is quite suitable for complex images where there are hard edges in the transition from background to foreground. In smooth image, MRF matting will partition the background and foreground colors into much less clusters. When ψ_F, ψ_B are both partitioned into only 1 cluster, MRF matting will be converted to perceptual matting.

Third, the results of our new method doesn't depend on a lot of hand-tweaking to select the right subregions and doesn't tune the parameters for each case. Our method runs automatically with the same settings for all cases and the sub-region segmentation doesn't matter. Sub-region partitioning is mainly for the purpose of decreasing the computational time.

6. Conclusions and Future Work

In this paper, we have presented a new approach to complex image matting using MRF model. First, the image is segmented into three regions, namely, foreground, background and unknown region. Then, the unknown region is segmented into several sub-regions. In each sub-region, the colors of background or foreground pixels are partitioned into clusters and the matting label is assigned to each unknown pixel in this subregion. The matting problem is modelled as a global optimization problem in MRF framework.

Our primary contribution is that we propose a novel complex image matting approach using MRF model. With a few more user interactions, this approach can obtain better results than previous natural image matting methods when matting on complex images. There are two main limitations remain in MRF matting. First, the estimated alpha in the label is still not accurate in every case; Second, the computing of likelihood and prior energy can be better modelled.

In the future, we hope to improve this method mainly from three aspects: alpha estimating ,energy computing and computational time. We also plan to implement an MRF matting system with adaptive influence parameters and use Graph-cut[14,15,16] to reducing the computational time.

References

[1] A. R. Smith and J. F. Blinn. Blue screen matting. In *Proceedings of SIGGRAPH 1996*, pages 259-268, August 1996.

[2] A. Berman, P. Vlahos, and A. Dadourian. Comprehensive method for removing from an image the back-ground surrounding a selected object. U.S.Patent 6,134,345, 2000.

[3] P. Hillman, J. Hannah and D. Renshaw. Alpha channel estimation in high-resolution images and image sequences. In *Proceedings of CVPR 2001*, pages 1063-1068, December 2001.

[4] M. Ruzon and C. Tomasi. Alpha estimation in natural images. In *Proceedings of CVPR 2000*, pages 18-25, June 2000.

[5] Y. Y. Chuang, B. Curless, D. Salesin and R. Szeliski. A Bayesian approach to digital matting. In *Proceedings of CVPR 2001*, pages 264-271, December 2001.

[6] J. Sun, J. Y. Jia, C. K. Tang, and H. Y. Shum. Poisson Matting. In *Proceedings of SIGGRAPH 2004*, pages 315-321, July 2004.

[7] Shengyou Lin, and Jiaoying Shi, Fast Natural Image Matting in Perceptual Color Space, *Computers and Graphics*, 29(3), pages 403-414, June 2005.

[8] S. Z. Li. Markov Random Field Modeling in Computer Vision. Springer-Verlag, 1995.

[9] G. Winkler. Random Fields and Dynamic Monte Carlo Methods. *Image analysis*, Springer Verlag, 1991.

[10] M. T. Orchard and C. A. Bouman. Color Quantization of Images. *IEEE Transaction on Signal Processing*, 39(12):2677-2690, December 1991.

[11] Y. Y. Chuang. Matting and Compositing. PhD thesis. 2004.

[12] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images. *IEEE Transaction on PAMI*, Vol. PAMI-6, pages 721-741, November, 1984.

[13] S. Kirkpatrick, C. D. Gellatt, Jr., and M. P. Vecchi. Optimization by Simulated Annealing. Science, Number 4598, May 1983.

[14] C. Rother, A. Blake, and V. Kolmogorov. Grabcut-Interactive Foreground Extraction Using Iterated Graph Cuts. In *Proceedings of SIGGRAPH 2004*, pages 309-314, July 2004.

[15] Y. Li, J. Sun, C. K. Tang, and H. Y. Shum. Lazy Snapping. In *Proceedings of SIGGRAPH 2004*, pages 303-308, July 2004.

[16] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive Image Segmentation Using an Adaptive GMMRF Model. In *Proceedings of ECCV 2004*, pages 428-442, May 2004.