Image Matting with Local and Nonlocal Smooth Priors

Xiaowu Chen[1], Dongqing Zou[1]*, Steven ZhiYing Zhou[2,3], Qinping Zhao[1], Ping Tan[2]
[1]State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University, Beijing, China
[2]Department of Electrical & Computer Engineering, National University of Singapore
[3]National University of Singapore (Suzhou) Research Institute

Abstract

In this paper we propose a novel alpha matting method with local and nonlocal smooth priors. We observe that the manifold preserving editing propagation [4] essentially introduced a nonlocal smooth prior on the alpha matte. This nonlocal smooth prior and the well known local smooth prior from matting Laplacian complement each other. So we combine them with a simple data term from color sampling in a graph model for nature image matting. Our method has a closed-form solution and can be solved efficiently. Compared with the state-of-the-art methods, our method produces more accurate results according to the evaluation on standard benchmark datasets.

1. Introduction

Image matting seeks to decompose an image I into the foreground F and the background B. Mathematically, the image I is a linear combination of F and B as the following:

$$C = F\alpha + B(1-\alpha). \quad (1)$$

Here, the alpha matte α defines the opacity of each pixel and its value lies in $[0, 1]$. Accurate matting plays an important role in various image and video editing applications. However, this problem is highly ill-posed, because the number of unknowns is much larger than that of equations in Equation (1). It is a typical practice to include a trimap or some user scribbles to simplify the problem. At the same time, strong priors on the alpha matte can significantly improve the results.

In the closed-form matting [11], a *matting Laplacian matrix* is derived based on the color line model [14] to constrain the alpha matte within local windows. This local smooth prior can be combined with data terms derived from color sampling [23, 15]. This smooth prior is further improved in [20] for image regions with constant foreground or background colors. He et al. [9] improve the color sampling with generalized Patchmatch [2]. Such a combination of data term and local smooth term generates high quality results according to recent surveys [22, 18]. However, as discussed in [10], it is nontrivial to set an appropriate local window size when computing the Laplacian matrix. A small window might not be sufficient to capture the structure details. On the other hand, a large window breaks the color line model and also leads to poor results.

Recently, Chen et al. [4] proposed a manifold preserving editing propagation method, and applied it to alpha matting. We observe that this method is essentially a novel nonlocal smooth prior on the alpha matte. It helps to correlate alpha values at faraway pixels, which is complementary to the matting Laplacian. When this nonlocal smooth prior is applied alone, it might not capture local structures of semi-transparent objects. So we propose to combine this nonlocal smooth term with the local Laplacian smooth term, and further include a trivial data term. This simple combination generates surprisingly good results as demonstrated by its performance on the benchmark database.

Our main contributions are: 1) new insights in the manifold preserving based matting propagation; 2) a novel matting algorithm that achieves superior performance on standard benchmark database.

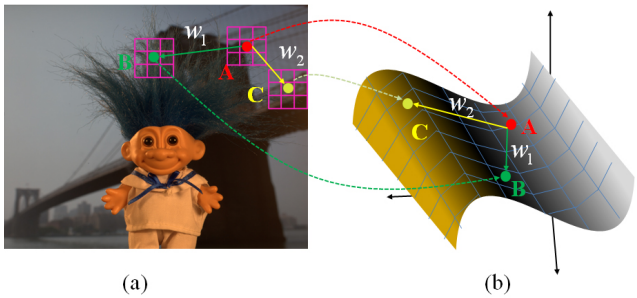
2. Related Work

Existing methods can be roughly categorized into sampling-based methods, affinity-based methods and combination of the two methods. We only discuss some of the most relevant works here. A more comprehensive survey can be found in [22].

Sampling-Based Matting estimates the alpha matte, foreground and background color of a pixel simultaneously. Many methods [16, 5, 15, 6, 3] applied different parametric or non-parametric models to collect nearby pixel samples from the known foreground and background. Ruzon and Tomasi [16] assumed the unknown pixels are in a narrow band region around the foreground boundary. Later

*corresponding author(email: zoudq@vrlab.buaa.edu.cn)

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.248



(a) (b)

Figure 1. (a) the input image. (b) the corresponding feature space. The pixel A can be generated by linearly combining the color at B and C.

this method was extended by Chuang et al. [5] with the Bayesian estimation framework. These methods work well when the unknown pixels are near the foreground boundary, and the number of unknown pixels is relatively small. Rhemann et al. [15] proposed an improved color model to collect samples according to the geodesic distance. Shared matting [6] collected those samples along rays of different directions. In general, these methods work well when the color is locally smooth.

Affinity-Based Matting solves the alpha matte independent of the foreground and background colors. Poisson matting [21] assumed that the matte gradient is proportional to the image gradient. Random walk matting [8] employed the random walks algorithm [7] to solve the alpha value according to the neighboring colors affinities. Closed-form matting [11] assumed color line model in local windows and solved the alpha matte by minimizing a cost function. Spectral matting [12] extended [11] into an unsupervised fashion by exploiting its relationship with spectral clustering. The matting Laplacian has been combined with various 'data constraints' [23, 15], priors [17] or learning based method [24] for image matting. However, the local smooth assumption is insufficient to deal with complex images. Thus we further combine it with a nonlocal smooth prior to improve the results.

Combined Sampling and Affinity Matting makes a good trade-off between the two approaches. Robust matting [23] first collected samples with high confidence, and then used the Random Walk [7] to minimize the matting energy. Global sampling matting [9] searched the global optimum samples with a random search algorithm derived from the PatchMatch algorithm [2].

3. Nonlocal Constraint on Alpha Matte

Chen et al. [4] proposed a manifold preserving method for edit propagation and applied it to propagate the alpha matte from the definite foreground and background to the unknown regions. Specifically, they first apply the locally linear embedding (LLE) [19] to represent each pixel as a linear combination of a few of its nearest neighbors in the RGBXY feature space, which is the RGB value concatenated with the image coordinate. The propagation algorithm keeps the alpha of known pixels unchanged, and requires every pixel to be the same linear combination of its neighbors in the feature space. For example, in the Figure 1, the pixel A can be generated by linearly combining the color at B and C according to the weight $A = w_1 B + (1 - w_1)C$. In the alpha matte, the manifold preserving propagation requires the same equation hold for the alpha matte, *i.e.* $\alpha_A = w_1 \alpha_B + (1 - w_1)\alpha_C$. Here, the three scalars α_A, α_B and α_C are the alpha values at the three pixels A, B and C. When B and C are known foreground and background pixels (*i.e.* $\alpha_B = 1$ and $\alpha_C = 0$), the manifold preserving condition simply requires $\alpha_A = w_1$. In fact, w_1 is an estimation of α_A from the color sampling [23]. So this method includes color sampling in the RGBXY space to solve the alpha. Furthermore, B and C can also be unknown pixels, so preserving the local manifold structure brings more constraints than color sampling. Note A could be faraway from B and C, since the neighbors are found in the feature space. So the manifold preserving constraint is essentially a nonlocal smooth constraint that relates the alpha values at faraway pixels. It propagates information across the whole image, while the local smooth constraint from matting Laplacian can only propagate information within local windows.

This nonlocal smooth prior is complementary to the matting Laplacian. As discussed in [10], it is sometimes difficult to choose an appropriate window size to compute the Laplacian matrix. As shown in the first row of Figure 2, a large window (41×41) is required to capture the alpha matte structure by the method [11]. However, for the example in the second row of Figure 2, a small window (3×3) is required to ensure the color line model to be valid over complicated background. He et al. [10] designed an adaptive method to choose different window sizes for different image regions. In comparison, when combined with the nonlocal smooth prior, matting Laplacian with a small window (3×3) generates consistently good results on both examples. Furthermore, a small window makes the matting Laplacian more sparse and more efficient to solve.

The nonlocal smooth prior alone is also insufficient to compute accurate alpha matte. As shown in the first row of Figure 3, when most of the alpha values are close to 0 or 1, the nonlocal smooth prior alone generates satisfactory results. However, as shown in the second row, the nonlocal smooth prior alone cannot handle images with large semi-transparent region. This is because for pixels with alpha values close to 0.5, their feature space neighbors are all unknown pixels. As a result, their alpha values are less constrained by the manifold preserving constraint. In contrast, when the local smooth constraint from matting Laplacian is applied, good results can be obtained on both examples.

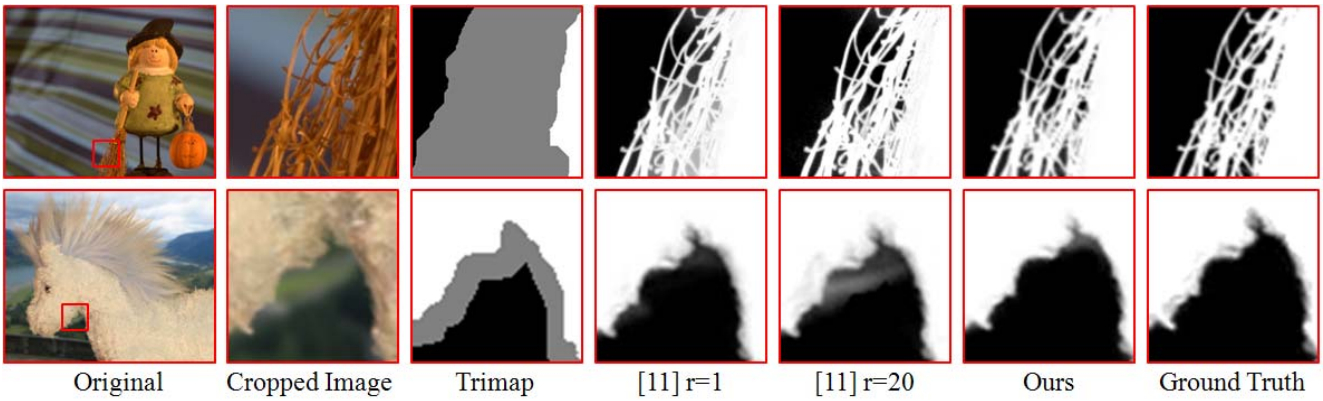


Figure 2. It is not easy to set the local window size when computing the matting Laplacian. In the first row, a large window (41×41) is required to capture the complex alpha matte by the method [11]. In the second row, a small window (3×3) is required to ensure the color line model on complex background. In comparison, when combined with nonlocal manifold prior, a small window (3×3) generates consistently good results on both examples.

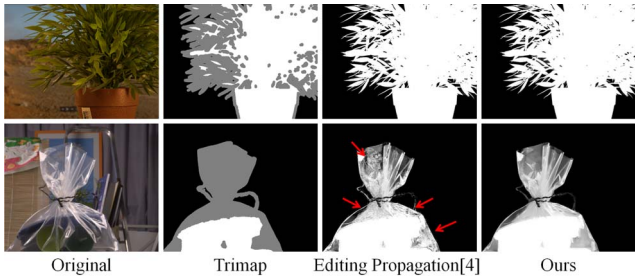


Figure 3. The nonlocal prior alone [4] is also insufficient to compute the accurate matte. It works for alpha matte with mostly binary alpha values. As shown in the second row, large semitransparent region causes artifacts.

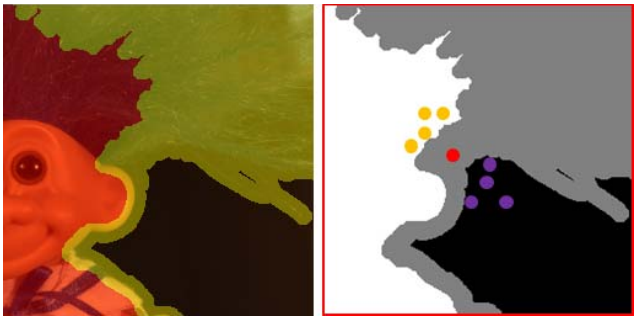


Figure 4. Color sampling. For the unknown pixel (the red point), we simply select the spatial nearest pixels as the candidate samples (orange points and purple points) by using FLANN [13].

4. Graph Model for Matting

We propose a new matting method by combining the local smooth term, nonlocal smooth term and a data term based on color sampling in a graph model. The result optimization can be solved efficiently from linear equations.

4.1. Color Sampling

We take a simple color sampling method [23] as the data term. Here we only briefly describe the basic idea, for more details please refer to [23]. Given a selected foreground and background samples pair (F_i, B_j), the alpha value of pixel C can be estimated as,

$$\widetilde{\alpha} = \frac{(C - B_j)(F_i - B_j)}{\|F_i - B_j\|^2}, \tag{2}$$

The distance $d(F_i, B_j)$ between the C and $\widetilde{\alpha}F_i + (1 - \widetilde{\alpha})B_j$ provides a confidence measure for the alpha value. The confidence is further modulated by the spatial distance between the samples and C. For the sampling, we simply take some spatial nearest pixels to form a candidate set, and choose the pair of samples with highest confidence for each unknown pixel as shown in Figure 4. We find the K nearest samples for each unknown pixel by using FLANN [13]. Though better sampling methods [23, 9] can be used, we find this simple method produces good results with the help of local and nonlocal smooth terms

4.2. Graph Model

As shown in Figure 5, in our graph model, the white nodes represent the unknown pixels on the image lattice, the orange nodes and the purple nodes are known pixels marked by a trimap or user scribbles. Two virtual nodes Ω_F and Ω_B representing the foreground and background are connected with each pixel. Each pixel is connected with its neighboring pixels in a 3×3 window, and also connected with its neighbors in the RGBXY feature space. The connections between each pixel and its feature space neighbors are indicated by red lines.

Data Term. The data weights, $W_{(i,F)}$ and $W_{(i,B)}$ which represent the probability of a pixel belonging to foreground and background, are defined between each pixel and a vir-

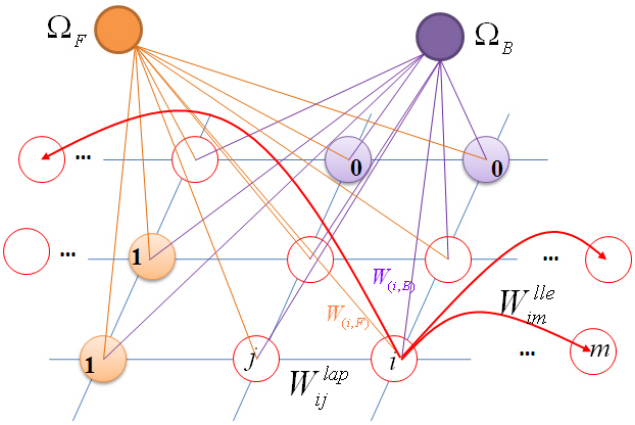


Figure 5. The graph model of our method. Two virtual notes representing the foreground and background are connected with each pixel. Each pixel is further connected with its spatial neighbors and feature space neighbors. Data term and smooth terms are defined on these edges.

tual node to enforce the data constraint. The initial alpha value $\widetilde{\alpha}$ of each pixel reflects the tendency towards foreground or background, thus it is reasonable to define the data weight of each node in the image lattice according to its initial alpha value. Specifically, for each unknown pixel i, two data weights $W_{(i,F)}$ and $W_{(i,B)}$ are defined as:

$$W_{(i,F)} = \gamma\widetilde{\alpha} \qquad W_{(i,B)} = \gamma(1-\widetilde{\alpha}).$$

The parameter γ balances the data term and the smooth term. We set γ to 0.1 in all our experiments. We use W_F and W_B to represent $\{W_{(i,F)}|i = 1,...,N\}$ and $\{W_{(i,B)}|i = 1,...,N\}$, respectively.

Local Smooth Term. The local matting Laplacian enhances the local smoothness of the result alpha matte. For the pixels i and j in a 3×3 window w_k, the neighbor term W_{ij}^{lap} is defined as:

$$W_{ij}^{lap} = \delta \sum_{k}^{(i,j)\in w_k} \frac{1 + (C_i - \mu_k)\left(\sum_k + \frac{\epsilon}{9}I\right)^{-1}(C_j - \mu_k)}{9}. \tag{3}$$

Here, the parameter δ controls the strength of the local smoothness. μ_k and $\sum_k$ represent the color mean and variance in each window. ϵ is a regularization coefficient which is set to 10^{-5}. Our window size is fixed as 3×3 for all examples.

Nonlocal Smooth Term. To enforce the nonlocal smooth constraint, for each pixel $\mathbf{X}_i$, we connect it to its K nearest neighbors $\mathbf{X}_{i1},...,\mathbf{X}_{ik}$ with weights W_{im}^{lle}. Specifically, the W_{im}^{lle} can be computed by minimizing

$$\sum_{i=1}^{N}\left\|\mathbf{X}_i - \sum_{m=1}^{K} W_{im}^{lle}\mathbf{X}_{im}\right\|^2, \tag{4}$$

subject to the constraint $\sum_{m=1}^{K} W_{im}^{lle} = 1$[19]. For our matting problem, the $\mathbf{X}_i$ represents the $(r_i, g_i, b_i, x_i, y_i)$ for the pixel i. (r_i, g_i, b_i) is its RGB color and (x_i, y_i) is its image coordinate. The result matrix $\mathbf{W}^{lle}$ (the ij-th element of $\mathbf{W}^{lle}$ is W_{ij}^{lle}) encodes the nonlocal manifold constraint.

4.3. Closed-form Solution

We first collect a subset of pixels $\mathcal{S}$, which includes pixels of known alpha value from the trimap or user scribbles. $\mathcal{S}$ also includes pixels whose initial alpha $\widetilde{\alpha}$ estimation is of high confidence (confidence larger than 0.85 in all our experiments). The energy function for solving alpha value is defined as:

$$E = \lambda\sum_{i\in\mathcal{S}}(\alpha_i - g_i)^2 + \sum_{i=1}^{N}\left(\sum_{j\in N_i} W_{ij}(\alpha_i - \alpha_j)\right)^2, \tag{5}$$

where N is the number of all nodes in the graph model, including all nodes in the image lattice plus two virtual nodes Ω_F and Ω_B. W_{ij} represents three kinds of weights , containing local smooth term W_{ij}^{lap}, nonlocal smooth term W_{ij}^{lle} and data term $W_{(i,F)}$ and $W_{(i,B)}$. The set N_i is the set of neighbors of the pixel i, including neighboring pixels in 3×3 window and those K nearest neighbors in the RGBXY space, and the two virtual nodes. Function 5 can be further written in a matrix form as

$$E = (\alpha - G)^T\Lambda(\alpha - G) + \alpha^T L^T L\alpha, \tag{6}$$

in which

$$L_{ij} = \begin{cases} W_{ii} & : \quad if \;\; i = j, \\ -W_{ij} & : \quad if \;\; i \; and \; j \; are \; neighbors, \\ 0 & : \quad otherwise, \end{cases} \tag{7}$$

Here, the weight $W_{ii} = \sum_{j\in N_i} W_{ij}$. Λ is a diagonal matrix, and G is a vector. Λ_{ii} is 1000 if $i \in \mathcal{S}$ and 0 otherwise. G_i is set to 1 if i belongs to foreground and 0 otherwise.

Equation 6 is a quadratic function about α, which can be minimized by solving the linear equation in closed-form solution

$$\left(\Lambda + L^T L\right)\alpha = \Lambda G. \tag{8}$$

5. Experiments

We evaluate the performance of our method and compare it with several recent representative methods. More results can be found in our supplement files.

5.1. Evaluation on Benchmark Database

We use the benchmark database [18] to evaluate the performance of our method[1]. Table 1 shows the ranking of

[1]Our results was uploaded to the evaluation web [1] at about 14:00 GMT November 15, 2012.

SAD	
Matting methods	Overall rank
1. Our	**4.2**
2. SVR	5.8
3. Weighted C&T	6.2
4. Shared	6.8
5. Global Sampling	8.2
6. Segmentation-based	8.5
7. Fast Automatic	8.5
8. Improved color	9.1
9. LSR	9.8
10. Global Sampling (F)	10

MSE	
Matting methods	Overall rank
1. Our	**3.5**
2. SVR	5.6
3. Weighted C&T	6.9
4. Global Sampling	7
5. Shared	7.6
6. KNN	8.5
7. Segmentation-based	8.8
8. Improved color	8.9
9. Fast Automatic	9.8
10. LSR	10.7

Gradient error	
Matting methods	Overall rank
1. Our	**4.5**
2. SVR	5.5
3. Shared	6.2
4. Segmentation-based	6.3
5. Global Sampling	6.4
6. Improved color	6.6
7. Weighted C&T	9.5
8. Global Sampling (F)	9.8
9. Fast Automatic	11
10. LMSPIR	11.3

Connectivity error	
Matting methods	Overall rank
1. Random Walk	3.2
2. Closed-Form	5.4
3. Our	**5.8**
4. Large Kernel	7.9
5. SVR	8.5
6. Improved color	8.8
7. LSR	9.6
8. Cell-based	9.9
9. Learning Based	10.1
10. Segmentation-based	11

Table 1. Ranks of different matting methods with respect to the four measurements on benchmark dataset as evaluated by [1]. We only show the top ten methods for each measurement.

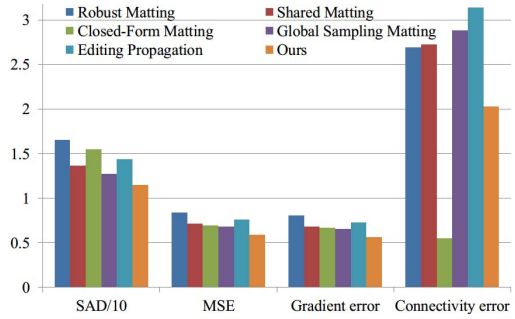


Figure 6. Quantitative evaluation of our method and the methods of Robust Matting [23], Shared Matting [6], Closed-Form Matting [11], Global Sampling Matting [9] and Editing Propagation [4].

different methods based on four measurements, *i.e.* absolute differences (SAD), mean squared error (MSE), gradient error and connectivity error. These results are generated from the website [1]. We only show the top ten methods for each measurement. Our method ranks first according to the measurements of SAD, MSE and gradient error, and ranks third according to the connectivity error. Although Random Walk Matting [8] and Closed-Form Matting [11] generate less connectivity error, they rank much lower according to the other three metrics..

We compare the quantitative error of our work with some recent methods (including Robust Matting [23], Shared Matting [6], Closed-Form Matting [11], Global Sampling Matting [9] and Editing Propagation [4]) in Figure 6. Our method generates smallest errors. The Global Sampling Matting has comparable accuracy by introducing a strong data term which optimizes color sampling from all pixel pairs. In comparison, our method works well with a much simpler data term. In fact, these two methods could be easily integrated for better performance, since they exploit different information. Several of these methods include the local smoothness term from matting Laplacian. However, as we discussed earlier, it is not easy to set a common window size for all test data. In comparison, with the help of nonlocal smoothness constraint, our method generates consistently good results with a small fixed window size.

5.2. Visual Comparison

We visually compare our method with Robust Matting [23], Shared Matting [6], Closed-Form Matting (CF) [11], Global Sampling Matting (GS) [9] and Editing Propagation (EP) [4] on difficult images. We selected three representative images. The foreground and background in the *Elephant* example have similar color distribution. The *Plasticbag* has large semi-transparent region, and the *Net* consists of many small holes. Figure 7 shows the results produced by our method and the other five methods. Visually, we obtain similar or better results. Our method also works with a scribble user interface. Some example results with sparse user scribbles and more results are included in our supplement files.

6. Conclusion

We observe that the manifold preserving alpha matte propagation is an effectively nonlocal smooth constraint on the alpha matte. We combine it with the local smooth constraint from the matting Laplacian and a simple data term from color sampling for nature image matting. Our graph model based method has closed-form solution and can be solved efficiently. Comparison on standard benchmark database shows our work outperforms state-of-the-art methods.

7. Acknowledgement

We want to thank Feng Ding and Jianwei Li for data processing. This work was partially supported by NSFC (60933006), 863 Program (2012AA011504) and ITER (2012GB102008). Ping Tan was supported by the Singpoare A*STAR PSF Project R-263-000-698-305. Steven Zhou receives financial support from the National University of Singapore (Suzhou) Research Institute.

References

[1] `http://www.alphamatting.com`.

[2] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.*, 28(3):24:1–24:11, July 2009.

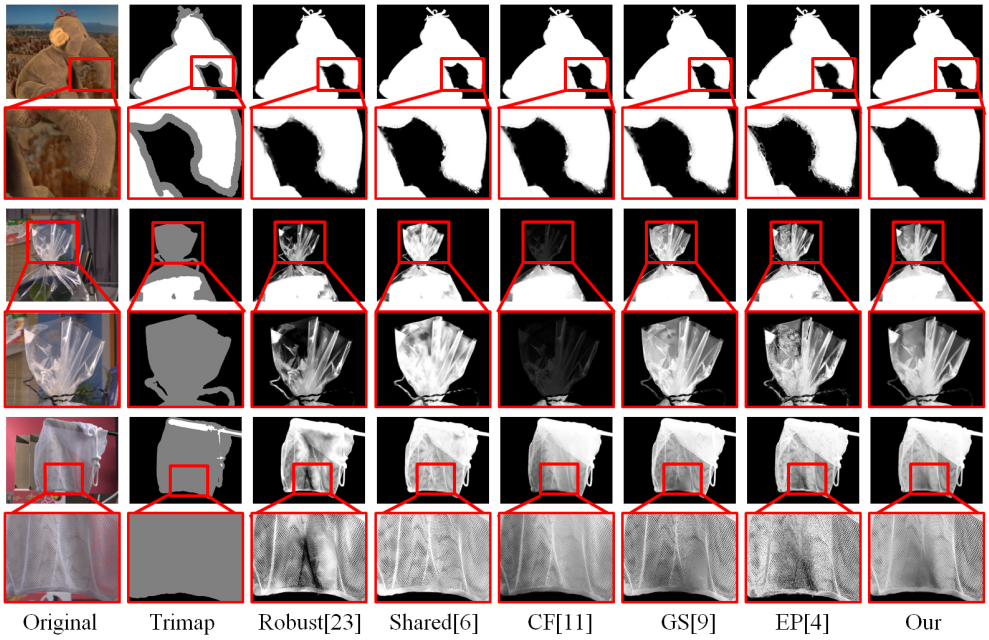


Figure 7. Visual comparison of our method with Robust Matting [23], Shared Matting [6], Closed-Form Matting (CF) [11], Global Sampling Matting (GS) [9] and Editing Propagation (EP) [4] on difficult images.

[3] X. Chen, Q. Li, Y. Song, X. Jin, and Q. Zhao. Supervised geodesic propagation for semantic label transfer. In *Proc. of ECCV*, pages 553–565, 2012.

[4] X. Chen, D. Zou, Q. Zhao, and P. Tan. Manifold preserving edit propagation. *ACM Trans. Graph.*, 31(6):132:1–132:7, Nov. 2012.

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

[6] E. S. L. Gastal and M. M. Oliveira. Shared sampling for real-time alpha matting. *Comput. Graph. Forum*, 29(2):575–584, 2010.

[7] L. Grady. Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(11), 2006.

[8] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann. Random walks for interactive alpha-matting. In *Proc. of International Conference on Visualization, Imaging and Image Processing*, pages 423–429, 2005.

[9] K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun. A global sampling method for alpha matting. In *Proc. of CVPR*, pages 2049–2056, 2011.

[10] K. He, J. Sun, and X. Tang. Fast matting using large kernel matting laplacian matrices. In *Proc. of CVPR*, pages 2165–2172, 2010.

[11] A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. In *Proc. of CVPR*, pages 61–68, 2006.

[12] A. Levin, A. Rav-Acha, and D. Lischinski. Spectral matting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(10):1699–1712, 2008.

[13] M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *Proc. of International Conference on Computer Vision Theory and Applications*, pages 331–340, 2009.

[14] I. Omer and M. Werman. Color lines: image specific color representation. In *Proc. of CVPR*, pages 946–953, 2004.

[15] C. Rhemann, C. Rother, and M. Gelautz. Improving color modeling for alpha matting. In *Proc. of British Machine Vision Conference*, 2008.

[16] C. Rhemann, C. Rother, P. Kohli, and M. Gelautz. Alpha estimation in natural images. In *Proc. of CVPR*, pages 1018–1025, 2000.

[17] C. Rhemann, C. Rother, A. Rav-Acha, and T. Sharp. High resolution matting via interactive trimap segmentation. In *Proc. of CVPR*, 2008.

[18] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott. A perceptually motivated online benchmark for image matting. In *Proc. of CVPR*, pages 1826–1833, 2009.

[19] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[20] D. Singaraju, C. Rother, and C. Rhemann. New appearance models for natural image matting. In *Proc. of CVPR*, pages 659–666, 2009.

[21] J. Sun, J. Jia, C.-K. Tang, and H.-Y. Shum. Poisson matting. *ACM Trans. Graph.*, 23(3):315–321, Aug. 2004.

[22] J. Wang and M. F. Cohen. Image and video matting: a survey. *Found. Trends. Comput. Graph. Vis.*, 3(2):97–175, Jan. 2007.

[23] J. Wang and M. F. Cohen. Optimized color sampling for robust matting. In *Proc. of CVPR*, 2007.

[24] Y. Zheng and C. Kambhamettu. Learning based digital matting. In *Proc.of ICCV*, pages 889–896, 2009.