Image Matting with Local and Nonlocal Smooth Priors

Xiaowu Chen\textsuperscript{1}, Dongqing Zou\textsuperscript{1,*}, Steven ZhiYing Zhou\textsuperscript{2,3}, Qinping Zhao\textsuperscript{1}, Ping Tan\textsuperscript{2}

\textsuperscript{1}State Key Laboratory of Virtual Reality Technology and Systems
School of Computer Science and Engineering, Beihang University, Beijing, China
\textsuperscript{2}Department of Electrical & Computer Engineering, National University of Singapore
\textsuperscript{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 \cite{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:

\begin{equation}
C = F\alpha + B(1-\alpha).
\end{equation}

Here, the alpha matte $\alpha$ 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 \cite{11}, a matting Laplacian matrix is derived based on the color line model \cite{14} to constrain the alpha matte within local windows. This local smooth prior can be combined with data terms derived from color sampling \cite{23, 15}. This smooth prior is further improved in \cite{20} for image regions with constant foreground or background colors. He et al. \cite{9} improve the color sampling with generalized Patchmatch \cite{2}. Such a combination of data term and local smooth term generates high quality results according to recent surveys \cite{22, 18}. However, as discussed in \cite{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. \cite{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 datasets.

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 \cite{22}.

**Sampling-Based Matting** estimates the alpha matte, foreground and background color of a pixel simultaneously. Many methods \cite{16, 15, 16, 3} applied different parametric or non-parametric models to collect nearby pixel samples from the known foreground and background. Ruzon and Tomasi \cite{16} assumed the unknown pixels are in a narrow band region around the foreground boundary. Later
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. Rhehmann 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 [14] 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 $C$.

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 \( \alpha_A, \alpha_B \) and \( \alpha_C \) 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 $\alpha_A$ 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 4 a large window ($41 \times 41$) is required to capture the alpha matte structure by the method [11]. However, for the example in the second row of Figure 4 a small window ($3 \times 3$) 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 ($3 \times 3$) 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 5 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 1. (a) the input image. (b) the corresponding feature space. The pixel $A$ can be generated by linearly combining the color at $B$ and $C$.](image-url)
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,

\[
\tilde{\alpha} = \frac{(C - B_j)(F_i - B_j)}{\|F_i - B_j\|^2}, \tag{2}
\]

The distance \(d(F_i, B_j)\) between the \(C\) and \(\tilde{\alpha}F_i + (1 - \tilde{\alpha})B_j\) 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 \(\Omega_F\) and \(\Omega_B\) representing the foreground and background are connected with each pixel. Each pixel is connected with its neighboring pixels in a \(3 \times 3\) 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-
subject to the constraint $\sum_{m=1}^{K} W_{im}^{lle} = 1$\cite{19}. For our matting problem, the $X_i$ represents the $(r_i, g_i, b_i, x_i, y_i)$ for the pixel $i$. $(r_j, g_j, b_j)$ is its RGB color and $(x_i, y_i)$ is its image coordinate. The result matrix $W^{lle}$ (the $ij$-th element of $W^{lle}$ is $W_{ij}^{lle}$) encodes the nonlocal manifold constraint.

4.3. Closed-form Solution

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

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

(5)

where $N$ is the number of all nodes in the graph model, including all nodes in the image lattice plus two virtual nodes $\Omega_F$ and $\Omega_B$. $W_{ij}$ 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 $3 \times 3$ 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

$$E = (\alpha - G)^T \Lambda (\alpha - G) + \alpha^T L^T L \alpha,$$

(6)

in which

$$L_{ij} = \begin{cases} W_{ij} & : \text{if } i = j, \\ -W_{ij} & : \text{if } i \text{ and } j \text{ are neighbors}, \\ 0 & : \text{otherwise}, \end{cases}$$

(7)

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

Equation [6] is a quadratic function about $\alpha$, which can be minimized by solving the linear equation in closed-form solution

$$(\Lambda + L^T L) \alpha = \Lambda G.$$

(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\textsuperscript{1}. Table [1] shows the ranking of

\textsuperscript{1}Our results was uploaded to the evaluation web [1] at about 14:00 GMT November 15, 2012.
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 Singapore A*STAR PSF Project R-263-000-698-305. Steven Zhou receives financial support from the National University of Singapore (Suzhou) Research Institute.

References

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.