LOCAL MATTING BASED ON SAMPLE-PAIR PROPAGATION AND ITERATIVE REFINEMENT

Bei He, Guijin Wang, Zhiwei Ruan, Xuanwu Yin, Xiaokang Pei, Xinggang Lin

Department of Electronic Engineering, Tsinghua University, China

ABSTRACT

This paper proposes a novel local matting algorithm based on sample-pair propagation and iterative refinement. Since sample-pairs of the foreground and background in the neighborhood are limited, they fail to fit the linear model well. We propose a sample-pair propagation scheme which propagates the confident sample-pair of each pixel to its neighbors so that they can collect more confident sample-pairs to estimate alpha values accurately. To avoid high time and space complexity of the global optimization, we convert matting into a de-noising problem and refine alpha values via fitting the linear model and smoothing the alpha matte locally and iteratively. Experimental results demonstrate that our algorithm produces more accurate results than the state-of-the-art of local matting.

Index Terms— local matting, gradient collection, sample-pair propagation, iterative refinement

1. INTRODUCTION

Matting refers to extracting the foreground from images/videos softly and accurately. It has been applied in lots of image/video editing tasks, such as layer separation, background replacement and foreground toning [1]. Specifically, the color of the pixel $\mathbf{I}_i$ in the input image can be modeled as a linear composite of the foreground $\mathbf{F}_i$ and background $\mathbf{B}_i$, where $\mathbf{I}_i{=}\alpha_i\mathbf{F}_i{+}(1{-}\alpha_i)\mathbf{B}_i$. α_i denotes the opacity of the foreground. Unknown parameters ($\mathbf{F}_i$, $\mathbf{B}_i$ and α_i) are more than equations, making matting an ill-posed problem. Therefore, a user-specified trimap is often required, where foreground, background and unknown pixels are marked.

Matting methods can mainly be categorized as sampling-based, affinity-based and a combination of the two [1]. The combination-based one utilizes the relevance of color values between the foreground/background samples and current pixel, while the affinity of the alpha matte is guaranteed. Thus, recent works fall into the combination-based category [2–6]. Robust Matting [3] firstly exploited non-parametric sampling to estimate alpha values, followed by the random walk optimization to refine matting results. Improved Color Matting [4] collected samples according to the geodesic distance [7], as opposed to the Euclidean distance [3]. However, those methods needed a global optimization to guarantee the affinity of the alpha matte which had high time and space complexity. In contrast, Real-Time Shared Matting [5] optimized the global energy function, but the drawback is that the linear model was not fitted well.

In this paper, we present a local matting algorithm based on sample-pair propagation and iterative refinement, named as LMSPIR. First, since unknown regions concentrate on the border between the foreground and background, we shoot rays in gradient directions to collect samples effectively. Second, confident sample-pairs of pixels are constructed by foreground/background samples and propagated to their neighbors. Sample-pair propagation helps to pick up more confident sample-pairs and estimate alpha values accurately. Third, we convert matting into a local de-noising problem to refine alpha values by fitting the linear model and smoothing the alpha matte iteratively. Experiments verify the accuracy of our LMSPIR algorithm compared with the state-of-the-art in local matting visually and quantitatively.

2. OUR LMSPIR ALGORITHM

Fig.1 illustrates a schematic of our LMSPIR algorithm. The output foreground is extracted from the input image and trimap. First, we collect foreground and background samples for each pixel via shooting rays in its gradient direction. Each combination of a foreground sample and a background one forms a sample-pair. Second, confident sample-pairs of pixels are propagated to their neighbors whose alpha values and corresponding confidences are updated. Finally, we refine alpha values via a local optimization which smoothes the alpha matte and fits the linear model iteratively.

2.1. Sample-pair Selection

2.1.1. Sample-pair Construction

According to the linear model, each pixel requires foreground and background samples to estimate the alpha value. In [3, 4], dense samples were collected nearby which introduced redundant samples. [5] selected sparse samples by shooting rays in random directions which failed to collect samples effectively. In contrast, we propose to collect samples by shooting rays

This paper is partially sponsored by NSFC (No. 61132007).

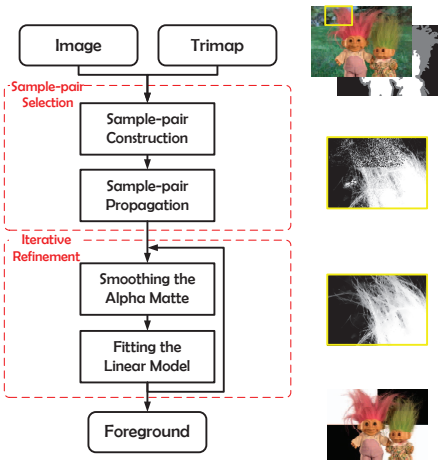

Fig. 1. A schematic of our LMSPIR algorithm.

in gradient directions. Since unknown regions concentrate on the border between the foreground and background, gradient directions help us to collect samples more effective than random ones. For edges (corresponding to high gradient), we shoot rays in gradient directions. As shown in Fig.2, gradient directions (marked as black dash lines) run through the foreground and background, in the direction where samples can be collected effectively. For smooth regions (corresponding to low gradient), we shoot rays in random directions which is similar to [5]. Shooting rays in different directions allows sample diversity and would provide more and better choices for neighboring pixels.

Rays in gradient lines spread in 2 directions when the closest foreground and background samples are collected in each direction. Hence at most 4 sample-pairs ($\mathbf{F}_i^{(m)}+\mathbf{B}_i^{(n)}$, $m,n\in\{1,2\}$) for each pixel are constructed. We preserve the most confident one among them to estimate the alpha value and confidence. The residual error of the linear model $r(\mathbf{F}_i^{(m)},\mathbf{B}_i^{(n)})$, the similarity to the foreground sample $s(\mathbf{F}_i^{(m)})$ and background one $s(\mathbf{B}_i^{(n)})$ are considered simultaneously, which are defined as,

$$\begin{aligned} r(\mathbf{F}_i^{(m)},\mathbf{B}_i^{(n)}) &= \frac{\|\mathbf{I}_i-\hat{\alpha}_i\mathbf{F}_i^{(m)}-(1-\hat{\alpha}_i)\mathbf{B}_i^{(n)}\|^2}{\|\mathbf{F}_i^{(m)}-\mathbf{B}_i^{(n)}\|^2} \\ s(\mathbf{F}_i^{(m)}) &= exp\{-T_{SC}^2/\|\mathbf{F}_i^{(m)}-\mathbf{I}_i\|^2\} \\ s(\mathbf{B}_i^{(n)}) &= exp\{-T_{SC}^2/\|\mathbf{B}_i^{(n)}-\mathbf{I}_i\|^2\}, \end{aligned} \tag{1}$$

where the alpha value $\hat{\alpha}_i$ is estimated by projecting the color of the current pixel onto the line spanned by the sample-pair [3]. T_{SC} refers to the threshold of the similarity between two color values. Combining those factors, we calculate a final confidence value for a sample pair as,

$$c(\mathbf{F}_i^{(m)},\mathbf{B}_i^{(n)}) = exp\{-\frac{r(\mathbf{F}_i^{(m)},\mathbf{B}_i^{(n)})\cdot s(\mathbf{F}_i^{(m)})\cdot s(\mathbf{B}_i^{(n)})}{\sigma_{SC}^2}\}, \tag{2}$$

Fig. 2. Samples are collected by shooting rays in gradient directions (marked as black dash lines).

where σ_{SC}^2 refers to the balance term.

2.1.2. *Sample-pair Propagation*

Samples collected by shooting rays are sparse, which are inadequate for each pixel. The reason is threefold. First, for pixels in smooth regions, sample-pairs could not be always found by shooting rays in random directions. Second, low confidence sample-pairs fail to fit the linear model which lead to inaccurate alpha values. Third, as alpha values of each pixel are estimated independently, the affinity of the alpha matte cannot be guaranteed. [5] exploited sample-pairs in the neighborhood of the pixel to pick up a more confident one; however, limited sample-pairs in the neighborhood cannot fit the linear model well. Our algorithm presents a scheme to propagate sample-pairs of pixels to neighbors. In this way, pixels can collect more confident sample-pairs to estimate accurate alpha values.

For pixels near the foreground/background, samples collected are more likely to be the true ones which lead to more confident alpha values. Additionally, pixels with low confidence sample-pairs tend to require more updates to estimate alpha values accurately. Therefore, our sample-pair propagation begins in pixels near the foreground/background and ends in ones far away. For the pixel i, sample-pair propagation includes two phases:

1. The collected sample-pair $(\mathbf{F}_i,\mathbf{B}_i)$ of the pixel i is propagated to pixels in the neighborhood $\mathcal{N}(i)$ with the window size w_{SP}.

2. The pixel $j\in\mathcal{N}(i)$ evaluates the confidence of the propagated sample-pair and preserves the more confident one compared with the current sample-pair $(\mathbf{F}_j,\mathbf{B}_j)$.

2.2. Iterative Refinement

2.2.1. *Refinement Criterion*

After collecting sample-pairs to initialize alpha values, inaccurate sample-pairs of the foreground and background lead to non-smooth regions of the alpha matte. As matting can be formulated as a Markov Random Field (MRF) model [1], we refine the alpha matte by minimizing the energy function

as $\min \sum_i (\|\alpha_i - \hat{\alpha}_i\|^2 + \lambda \sum_j \|\alpha_i - \alpha_j\|^2)$, $j \in \mathcal{N}(i)$ where λ refers to the balance between the data term and smoothness one. In [4], the matting Laplacian matrix [8] was introduced to optimize the energy function globally. Due to high time and space complexity of the matting Laplacian matrix, we novelly formulate it as a de-noising problem where the data term and smoothness one are minimized iteratively.

Our refinement criterion is composed of two parts. On one hand, assuming the error of the estimated alpha value $\hat{\alpha}_i$ of the pixel i is n_i, the linear model can be replaced by $\mathbf{I}_i = (\mathbf{F}_i - \mathbf{B}_i)(\hat{\alpha}_i + n_i) + \mathbf{B}_i$. Hence the data term can be minimized by reducing the estimated error n_i. On the other hand, we propose to estimate the alpha value $\hat{\alpha}_i$, taking neighboring ones into consideration where $\hat{\alpha}_i = \sum_j \hat{\alpha}_j / N$. N refers to the number of pixels in the neighborhood. Thus, the affinity of the alpha matte is guaranteed so that the smoothness term can be minimized. Via reducing the estimated error and smoothing the alpha matte alternately, we minimize the energy function of the MRF model locally.

2.2.2. *Iterative Optimization*

As color $\mathbf{I}_i$ can be regarded as a linear observation of the alpha value α_i, minimizing the estimated error is equivalent to minimizing the observation error. Additionally, smoothing the alpha matte can be regarded as estimating the new alpha value $\hat{\alpha}_i$ with the old $\hat{\alpha}_i$ and impulse of neighbors $\hat{\alpha}_j$. Since the above two conditions fit observation and state equations of the kalman filter respectively, we introduce the kalman filter to reduce the observation error and smooth the alpha matte iteratively and locally. The smoothing and reducing steps alternate until the decline rate of average alpha value variance is less than T_{IR}. In our experiments, the number of our iteration ranges from 5 to 10.

i) Smoothing the alpha matte

Since pixels with more similar color values indicate closer alpha values, color similarity is applied to evaluate the relevance of alpha values. The alpha value of the pixel i in $(k+1)^{th}$ iteration can be estimated by its neighbors,

$$\hat{\alpha}_i^{k+1|k} = \sum \omega_{ij}^{k|k} \hat{\alpha}_j^{k|k} / \sum \omega_{ij}^{k|k}, \tag{3}$$

where $\omega_{ij}^{k|k} = c_j \cdot exp\{-\|\mathbf{I}_i - \mathbf{I}_j\|^2 / \sigma_{IR}^2\}$ refers to color similarity of pixels i and j. $j \in \mathcal{N}(i)$ with the window size w_{IR}. σ_{IR} refers to the similarity threshold between two color values. Assuming each alpha value is an independent random variable, the variance of the alpha value is denoted as,

$$p_i^{k+1|k} = (\sum (\omega_{ij}^{k|k} (\hat{\alpha}_j^{k|k} - \hat{\alpha}_i^{k+1|k}))^2) / (\sum \omega_{ij}^{k|k})^2. \tag{4}$$

ii) Reducing the observation error

The alpha value of pixel i in $(k+1)^{th}$ iteration is updated by,

$$\hat{\alpha}_i^{k+1|k+1} = \hat{\alpha}_i^{k+1|k} + \mathbf{k}_i^{k+1|k^T} (\mathbf{I}_i - (\mathbf{F}_i - \mathbf{B}_i)\hat{\alpha}_i^{k+1|k} - \mathbf{B}_i). \tag{5}$$

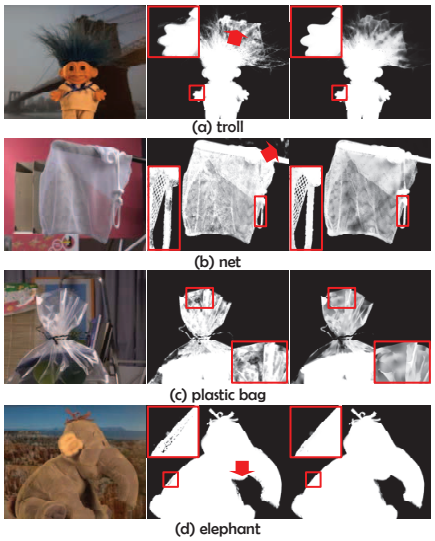

Fig. 3. Matting results of Real-Time Shared Matting [5] and our LMSPIR algorithm in [9].

The corresponding gain and variance of the alpha value are denoted as,

$$\mathbf{k}_i^{k+1|k^T} = \frac{p_i^{k+1|k}(\mathbf{F}_i - \mathbf{B}_i)^T}{p_i^{k+1|k}(\mathbf{F}_i - \mathbf{B}_i)(\mathbf{F}_i - \mathbf{B}_i)^T + \mathbf{R}_{IR}} \tag{6}$$
$$p_i^{k+1|k+1} = p_i^{k+1|k}[1 - \mathbf{k}_i^{k+1|k^T}(\mathbf{F}_i - \mathbf{B}_i)],$$

where $\mathbf{R}_{IR}$ refers to the variance of the observation error.

3. EXPERIMENTS AND DISCUSSION

To evaluate the performance of our LMSPIR algorithm, we implement and run it on the Intel Core II CPU with 2.0 GHz Dual. Parameter values are fixed as follows: $\sigma_{SC} = 0.1$, $T_{SC} = 15$, $w_{SP} = 15$, $T_{IR} = 0.1$, $\sigma_{IR} = 10$, $w_{IR} = 7$ and $R_{IR} = 0.2$. We use the test dataset provided by [9] that has been used in recent performance evaluation of matting. Our algorithm compares with Real-Time Shared Matting (RTSM) [5] which ranked top in local matting techniques [9].

3.1. Visual and Quantitative Comparison

Fig.3 shows several matting results of RTSM [5] and our LMSPIR algorithm. For each image, we list the original one (left), matting results of RTSM (middle) and our algorithm (right) respectively. For the "troll" image, we reduce the interference of the bridge in the background (marked by the red arrow) and achieve more smooth results in the region of the hand. For the "net" image, the stick affects alpha values of the background in RTSM (marked by the red arrow). Additionally, our results of the net and rope are more accurate than RTSM. For the "plastic bag" image, matting results of

Table 1. The average rankings of test images and overall rankings among all algorithms (in brackets) of the state-of-the-art local matting algorithms and our LMSPIR in [9].

Algorithm	SAD	MSE
LMSPIR	5.0 (4^{th})	5.5 (4^{th})
Real-time shared matting [5]	5.3 (5^{th})	6.0 (7^{th})
Robust matting [3]	7.8 (9^{th})	7.3 (9^{th})
Geodesic matting [7]	12 (12^{th})	12.3 (13^{th})
Iterative BP matting [2]	12.5 (13^{th})	11.4 (11^{th})

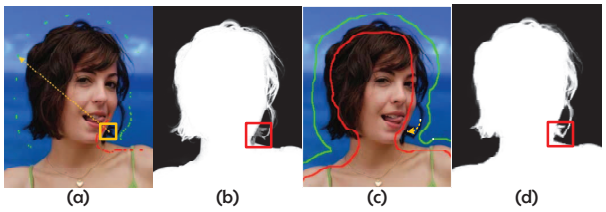

Fig. 4. Sample-pair selection results of Real-Time Shared Matting [5] and our LMSPIR algorithm.

ours are much more smooth than RTSM on the opening of the plastic bag. For the "elephant" image, the back of the elephant in our results is more smooth than RTSM. Further, regions of the nose marked by the red arrow in our results are more like the the original one.

The average rankings of the state-of-the-art local matting algorithms and our LMSPIR in SAD (Sum of Absolute Difference) and MSE (Mean Squared Error) evaluation for test images of [9] are illustrated in Table.1. [9] lists top 16 matting algorithms where our algorithm ranks both 4 in SAD and MSE (in brackets). The top 3 matting techniques in [9] all required a global optimization of the matting Laplacian matrix which had high time and space complexity. In conclusion, our algorithm provides more accurate results than other local matting methods both visually and quantitatively.

3.2. Module Analysis

Sample-pair selection results of RTSM and our LMSPIR algorithm are shown in Fig.4. For the pixel close to the tongue (marked as the white point in Fig.4(a)), RTSM selected the most confident sample-pair in the neighborhood (marked as the yellow rectangle); however, those samples of the foreground all concentrated on the face which can't fit the linear model well. As a result, RTSM reported inaccurate matting results in Fig.4(b). Our algorithm propagates sample-pairs of the pixel far away to the current one so that the correct sample of the foreground, i.e. the black hair, is picked up. As a result, we achieve more accurate alpha matte in Fig.4(d), shown by the red rectangle. Simultaneously, we take $7.3s$ for sample-pair selection, which is competitive with RTSM according to experiments in [6].

To refine alpha values, the matting Laplacian matrix optimization requires $O(n^2)$ space complexity for n unknown pixels in the trimap. But we only cost $O(n)$ space to save several temporary images. Additionally, our algorithm averagely costs $1.17sec$ on iterative refinement, which is faster than $3.19sec$ of the matting Laplacian matrix optimization.

4. CONCLUSION

This paper presents a local matting algorithm based on sample-pair propagation and iterative refinement. We propose a scheme which propagates confident sample-pairs of the foreground and background in the neighborhood. These sample-pairs help pixels to fit the linear model so that accurate matting results can be achieved. Furthermore, our algorithm converts matting into a de-noising problem where alpha values is refined by fitting the linear model and smoothing the alpha matte iteratively. Our local refinement avoids high time and space complexity of the global optimization. Visual and quantitative results prove our algorithm outperforms other local matting techniques.

References

[1] J. Wang and M.F. Cohen, "Image and video matting: a survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 2, pp. 97–175, 2007.

[2] J. Wang and M.F. Cohen, "An iterative optimization approach for unified image segmentation and matting," in *Proc. of ICCV*, 2005, vol. 2, pp. 936–943.

[3] J. Wang and M.F. Cohen, "Optimized color sampling for robust matting," in *Proc. of CVPR*, 2007, pp. 1–8.

[4] C. Rhemann, C. Rother, and M. Gelautz, "Improving color modeling for alpha matting," in *Proc. of BMVC*, 2008.

[5] E.S.L. Gastal and M.M. Oliveira, "Shared sampling for real-time alpha matting," in *Proc. of Computer Graphics Forum*, 2010, vol. 29, pp. 575–584.

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

[7] X. Bai and G. Sapiro, "A geodesic framework for fast interactive image and video segmentation and matting," in *Proc. of ICCV*, 2007, pp. 1–8.

[8] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 228–242, 2008.

[9] 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*, 2009, pp. 1826–1833.