PAPER

High-accuracy and Quick Matting based on Sample-pair Refinement and Local Optimization

Bei HE[†a)], ***Student Member***, **Guijin WANG**[†b)], ***Member***, **Chenbo SHI**[†], ***Student Member***, **Xuanwu YIN**[†], **Bo LIU**[†], ***and*** **Xinggang LIN**[†], ***Nonmembers***

SUMMARY Based on sample-pair refinement and local optimization, this paper proposes a high-accuracy and quick matting algorithm. First, in order to gather foreground/background samples effectively, we shoot rays in hybrid (gradient and uniform) directions. This strategy utilizes the prior knowledge to adjust the directions for effective searching. Second, we refine sample-pairs of pixels by taking into account neighbors'. Both high confidence sample-pairs and usable foreground/background components are utilized and thus more accurate and smoother matting results are achieved. Third, to reduce the computational cost of sample-pair selection in coarse matting, this paper proposes an adaptive sample clustering approach. Most redundant samples are eliminated adaptively, where the computational cost decreases significantly. Finally, we convert fine matting into a de-noising problem, which is optimized by minimizing the observation and state errors iteratively and locally. This leads to less space and time complexity compared with global optimization. Experiments demonstrate that we outperform other state-of-the-art methods in local matting both on accuracy and efficiency.

key words: *quick matting, hybrid direction, sample-pair refinement, adaptive sample clustering, local optimization, kalman filter*

1. Introduction

Matting, which aims to extract the foreground from images softly and accurately, has been applied in image/video editing applications, such as layer separation, foreground toning and background replacement [1]. In the matting problem, the color value $\mathbf{I}_i$ can be modeled as a linear composite of the foreground component $\mathbf{F}_i$ and the background one $\mathbf{B}_i$ for the pixel p_i:

$$\mathbf{I}_i = \alpha_i \mathbf{F}_i + (1 - \alpha_i)\mathbf{B}_i, \tag{1}$$

where $\alpha_i \in [0, 1]$ refers to the opacity of the foreground component and constitutes the alpha matte $\boldsymbol{\alpha}$. Since $\mathbf{F}_i$, $\mathbf{B}_i$ and α_i are unknown, matting is inherently an ill-posed problem. To achieve a semantically meaningful alpha matte [1], almost all matting approaches start by dividing the input image into 3 regions: definite foreground, definite background and unknown (named as the trimap).

Related matting methods can be categorized into sampling-based [2–4], affinity-based [5–11] and combinations of the two [12–17]. 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. Consequently, recent works fall into the combination-based category [12–17]. Wang [13] firstly exploited the color-sampling strategy to estimate alpha values, followed by the random walk optimization [18] to smooth matting results. Opposed to the Euclidean distance [13], Rhemann [14] introduced the geodesic distance to gather samples. Additionally, the alpha matte was formulated as a Markov Random Field (MRF) model and optimized globally. Considering collected sample-pairs were insufficient, Gastal [15] refined sample-pairs of pixels with ones in the neighborhood. He [16] extended the search of samples to the whole image so that a more accurate alpha matte could be calculated. However, the corresponding computational cost was heavy. On one hand, collecting and selecting the adequate sample-pair incurred large computation. On the other hand, global optimization led to more precise and smoother results, but the time and space complexity was unaffordable. Hence, high-accuracy and quick matting results cannot be achieved.

To overcome problems listed above, we provided a local matting method which refined sample-pairs in a propagation mode and optimized the alpha matte with the kalman filter in our earlier work [17]. However, this paper further expands the original idea and proposes an algorithm based on sample-pair refinement and local optimization. The new method improves the accuracy and efficiency of matting signicantly. Our contributions are summarized as follows. *1)* We shoot rays in hybrid (gradient and uniform) directions to collect foreground/background samples. Prior knowledge is utilized adequately so sample-pairs of pixels can be constructed effectively. *2)* For high-accuracy alpha values of pixels, we propose to refine their sample-pairs by considering neighboring ones. First, both high-confidence sample-pairs and usable foreground/background components are shared in the neighborhood. Second, pixels are refined from high confidence to low and thus low confidence pixels can obtain better sample-pairs for refinement. *3)* To speed up the coarse matting step significantly, this paper presents an adaptive sample clustering strategy for sample-pair selection. On one hand, most redundant samples are eliminated by sample clustering. On the other hand, the number of sample-pairs for each pixel can be determined independently and adaptively. *4)* We novelly convert fine matting into a de-noising problem. The alpha matte is refined by minimizing the observation and state errors iteratively, where expensive time and space complexity of global optimization

Manuscript received January 1, 2010.
Manuscript revised January 1, 2010.
[†]The authors are with the the Dept. of Electronic Engineering, Tsinghua University, ROC
a) E-mail: b-he08@mails.tsinghua.edu.cn
b) E-mail: wangguijin@tsinghua.edu.cn
DOI: 10.1587/transinf.E0.D.1

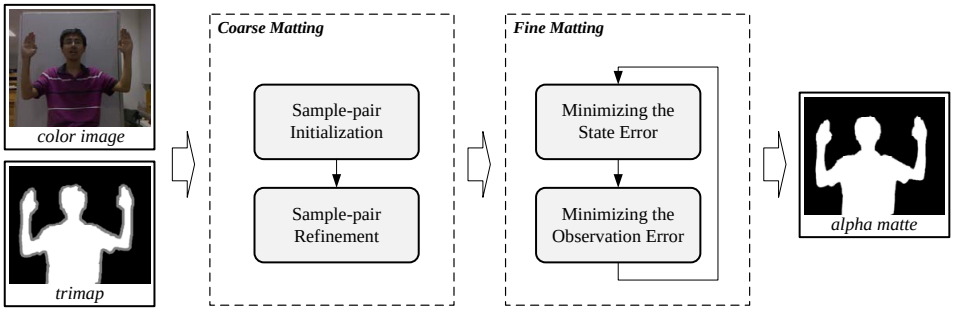


Fig. 1 Overview of our high-accuracy and quick matting algorithm.

is avoided.

The rest of the paper is organized as follows. In Section 2, we review the related work and put forward our motivations. Details of our coarse-to-fine matting strategy are presented from Section 3 to Section 5 respectively. We show experiments and discussions in Section 6, followed by a brief conclusion in Section 7.

2. Related Work and Motivation

A brief overview of our algorithm is illustrated in Fig.1. First, the color image and trimap are prepared. Second, the coarse alpha matte is estimated from collected sample-pairs via initialization and refinement. Third, we perform a local optimization for fine matting, which minimizes the state and observation errors iteratively. In this section, we mainly review the related work on above modules, which are relevant to our contributions.

1) Sample-pair Initialization: Recent works [13–17] gathered multiple pixels in definite foreground/background regions as samples to initialize the alpha matte, named as color-sampling. Wang [13] collected dense samples on the basis of the minimal Euclidean distance. Opposed to that, Rhemann [14] collected samples according to the minimal geodesic distance [8]. However, adjacent samples have similar color values which are redundant to estimate alpha values. Gastal [15] shot sparse rays in uniform directions to exploit samples. Since the prior knowledge are not utilized, isotropical directions are inadequate for edge pixels. On the contrary, our algorithm proposes to shoot rays in hybrid (gradient and uniform) directions, where pixels can gather samples effectively.

2) Sample-pair Refinement: Due to the limitation of computational cost, insufficient sample-pairs are constructed. Thus, inaccurate estimations are achieved by sample-pair initialization [13, 14]. Gastal [15] selected the most confident sample-pairs from ones in the neighborhood to refine the current sample-pair. In contrast, He [17] propagated sample-pairs among neighboring pixels. However, they neglected the fact that the foreground/background component of low confidence sample-pair may be useful. In our algorithm, both the high confidence sample-pairs and usable foreground/background components are employed, where more accurate matting results are provided. Additionally, we refine pixels according to the decreasing order of their confidences. Thus, low confidence pixels will gather more usable sample-pairs for refinement.

3) Fine Matting: Since the alpha matte is achieved pixel-by-pixel during color-sampling, there exist many non-smooth regions. Previous works [13, 14] formulated the fine matting as a MRF model and then applied the matting Laplacian matrix [7] to optimize globally. Considering the corresponding time and space complexity is unaffordable, we convert fine matting into a de-noising problem. Then we can achieve smooth alpha matte via minimizing the state and observation errors iteratively and locally.

3. Coarse Matting

This section describes how we estimate the alpha value, foreground and background components for each pixel via color-sampling. In the literature, foreground/background samples were collected and utilized to estimate the alpha value as,

$$\hat{\alpha}_i = (\mathbf{I}_i - \mathbf{B}_i)^T(\mathbf{F}_i - \mathbf{B}_i)/\|\mathbf{F}_i - \mathbf{B}_i\|^2. \tag{2}$$

Denote a foreground sample and a background one as a sample-pair. The color-sampling approach corresponds to collecting and selecting the most confident sample-pair for each pixel. To accomplish accurate and quick coarse matting, our method comprises 2 steps: sample-pair initialization and refinement. The first step aims to initialize sample-pairs of unknown pixels via shooting hybrid (gradient and uniform) directions for collection. In the following, we refine the sample-pair of each pixel, making use of its neighbors'.

3.1 Confidence Definitions

To select the most confident sample-pair, we take into account 3 factors: the residual error of the linear model cr_i, the dissimilarity with the foreground sample cf_i and the dissimilarity with the background one cb_i. Similar to our previous work [17], they are defined as,

$$\begin{aligned} cr_i &= \|\mathbf{I}_i - \hat{\alpha}_i\mathbf{F}_i - (1-\hat{\alpha}_i)\mathbf{B}_i\|^2/\|\mathbf{F}_i - \mathbf{B}_i\|^2 \\ cf_i &= \exp\{-I_{max}^2/\|\mathbf{F}_i - \mathbf{I}_i\|^2\} \\ cb_i &= \exp\{-I_{max}^2/\|\mathbf{B}_i - \mathbf{I}_i\|^2\} \end{aligned} \quad , \tag{3}$$

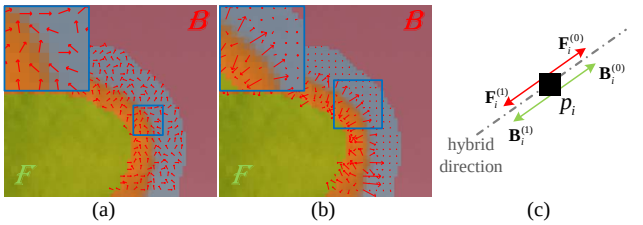


Fig. 2 Samples are gathered by shooting rays in uniform (a) and gradient (b) directions. (c) refers to an example to collect foreground and background samples.

where I_{max} is set to balance the color values.

1) The confidence of a sample-pair: The confidence of a sample-pair is achieved by combining the 3 factors above:

$$c_i = \exp\{-cr_i \cdot cf_i \cdot cb_i/\sigma_c^2\}, \tag{4}$$

where σ_c is set to balance the confidence value.

2) The confidence of a pixel: Among all collected sample-pairs of a pixel, only the one with the highest confidence is preserved. The corresponding confidence denotes the confidence of that pixel.

3.2 Sample-pair Initialization

For pixels in smooth regions, as no prior knowledge is available, shooting rays isotropically is adequate to exploit samples, e.g., in uniform directions [15]. However, gradient directions run through the foreground/background directly for pixels in edge regions, since unknown regions concentrate on the border between the foreground/background. Under the above condition, shooting rays in gradient directions can collect samples more effectively than uniform ones. Consequently, our algorithm gathers samples by shooting rays in hybrid directions. That is, gradient directions are employed for pixels in edge regions and uniform ones for pixels in smooth regions. The angle of the direction for the pixel p_i is formulated as,

$$\theta_i = \begin{cases} \theta_i^{grad}, if\ A_i^{grad} > T_a \\ \theta_i^{uni}, otherwise, \end{cases} \tag{5}$$

where θ_i^{grad} and θ_i^{uni} denote the angles of gradient and uniform directions respectively. A_i^{grad} represents the amplitude of the gradient and T_a is the corresponding threshold. An example of θ_i^{uni} and θ_i^{grad} is declared in Fig.2. θ_i^{uni} is randomly and uniformly selected from the set $\{0°, 20°, ..., 140°, 160°\}$.

Taking the collection of foreground samples as an examples, we shoot rays along directions with angles θ_i and θ_i+180^o respectively for each pixel, where the nearest pixels in definite foreground regions are gathered. Thus, at most 2 foreground samples are searched. The collection for background samples is similar. In Fig.2(c), an example of the sample collection is provided. Hence, we can construct at most 4 sample-pairs $(\mathbf{F}_i^{(m)}, \mathbf{B}_i^{(n)})$, m, n=0, 1 via crossed combination for the unknown pixel p_i. Among them, only the one with the highest confidence is preserved to initialize the alpha and confidence values.

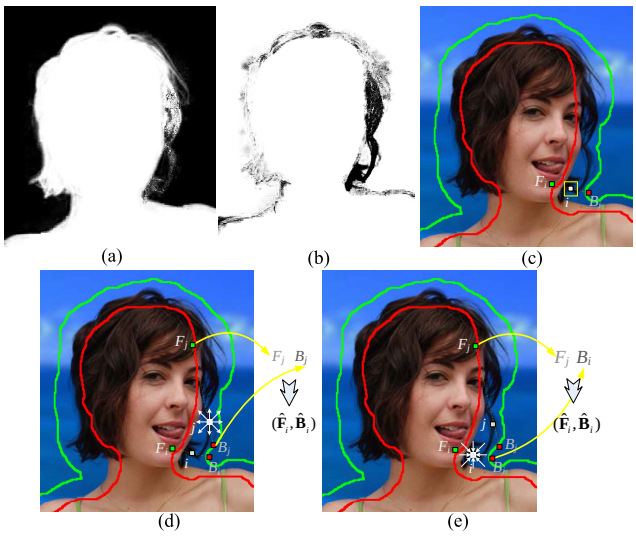


Fig. 3 $\mathbf{F}_i$: "yellow", $\mathbf{B}_i$: "dark blue", $\mathbf{F}_j$: "dark brown" and $\mathbf{B}_j$: "light blue". (a)-(b) refer to the initialized alpha and confidence mattes. Gastal's [15], He's [17] and our algorithm are listed in (c)-(e) respectively.

3.3 Sample-pair Refinement

Shooting rays help each pixel to preserve one sample-pair, but it is still sparse and inadequate for accurate estimations. We suggest refining sample-pairs of unknown pixels by considering both high confidence sample-pairs and usable foreground/background samples in the neighborhood. For the unknown pixel p_i, our refinement is arranged in 3 parts. First, the pixel collects N sample-pairs $(\mathbf{F}_j, \mathbf{B}_j)$ of N neighboring pixels p_j, $p_j \in \mathcal{N}(p_i)$. Second, those sample-pairs are split to N foreground/background samples and then reconstructed to new $N \times N$ sample-pairs. Finally, the sample-pair of highest confidence is preserved as new $(\mathbf{F}_i, \mathbf{B}_i)$ to refine the alpha value. Additionally, we refine unknown pixels from high confidence to low. Due to the decreasing order, low confidence pixels will be processed with refined sample-pairs of high confidence neighbors. It implies that more usable sample-pairs would be gathered during the refinement of low confidence pixels.

As shown in Fig.3(b), the initialized sample-pair $(\mathbf{F}_i, \mathbf{B}_i)$ is low confidence for the pixel p_i. Since in the limited neighborhood, Gastal [15] cannot exploit any high confidence sample-pair, inaccurate estimation is induced. Opposed to that, He [17] propagates the sample-pair of the pixel p_j to the current one, but actually only the foreground component is needed. Applying $(\mathbf{F}_j, \mathbf{B}_j)$ directly is not suitable if the background is gradient. In our algorithm, sample-pairs of pixels p_i and p_j are split and re-constructed to $(\mathbf{F}_i, \mathbf{B}_i)$, $(\mathbf{F}_i, \mathbf{B}_j)$, $(\mathbf{F}_j, \mathbf{B}_i)$ and $(\mathbf{F}_j, \mathbf{B}_j)$. Consequently, the pixel p_i can only employ the foreground component of the neighboring pixel p_j to form a higher confidence sample-pair $(\mathbf{F}_j, \mathbf{B}_i)$ and thus more accurate results will be achieved. The examples of the three algorithm are listed in Fig.3(c)-(e)

respectively.

The overall coarse matting, including sample-pair initialization and refinement, is summarized in Algorithm.1.

Algorithm 1 Coarse Matting

Sample-pairs Initialization:

1: unknown pixels are pushed into a queue **Q**.
2: **while Q** is not empty **do**
3: pop the first pixel $p_i \in \mathbf{Q}$;
4: *generate samples*: shoot rays in hybrid (gradient and uniform) directions to gather pixels in foreground/background regions as samples;
5: *generate sample-pairs*: at most 4 sample-pairs are constructed;
6: the highest confidence sample-pair is selected as $(\mathbf{F}_i, \mathbf{B}_i)$ to initialize alpha and confidence values of p_i;
7: **end while**
8: return initialized sample-pairs, alpha and confidence values of unknown pixels.

Sample-pairs Refinement:

9: unknown pixels are sorted by confidence values in descending order and pushed into a queue $\mathbf{Q}'$.
10: **while** $\mathbf{Q}'$ is not empty **do**
11: pop the first pixel $p_i \in \mathbf{Q}'$;
12: *generate samples*: gather N sample-pairs from neighboring unknown pixels p_j and split to N foreground and background samples respectively;
13: *generate sample-pairs*: re-construct those samples to $N \times N$ sample-pairs;
14: the highest confidence sample-pair is selected as new $(\mathbf{F}_i, \mathbf{B}_i)$ to refine alpha and confidence values of p_i;
15: **end while**
16: return refined sample-pairs, alpha and confidence values of unknown pixels.

4. Speedup of Coarse Matting

Recently, selecting the most confident sample-pair from $N \times N$ ones in traditional coarse matting confronts expensive computational cost [13–15, 17]. Hence, this section is devoted to reducing redundant sample-pairs for speedup particularly. Usually, high confidence pixels require less sample-pairs to refine alpha values compared with low confidence ones. If N is small, pixels cannot calculate alpha values precisely without enough sample-pairs; however, large N leads to expensive computational cost. Moreover, color redundance usually exists in collected samples of neighboring pixels due to the consistency of the image, especially in smooth regions.

An adaptive sample clustering strategy is designed to decide the value of N and reduce the redundance of samples. As shown in Fig.4, low confidence pixels need more samples to achieve lower residual errors. Therefore, we apply confidence values for the adaptive number N_i of the pixel p_i as,

$$N_i = N_{min} + (N_{max} - N_{min})(1 - c_i), \tag{6}$$

where N_{min} and N_{max} limit the range of N_i.

Then, we cluster the N_i foreground/background samples for the pixel p_i to eliminate the color redundancy. Since color values between samples of neighboring pixels

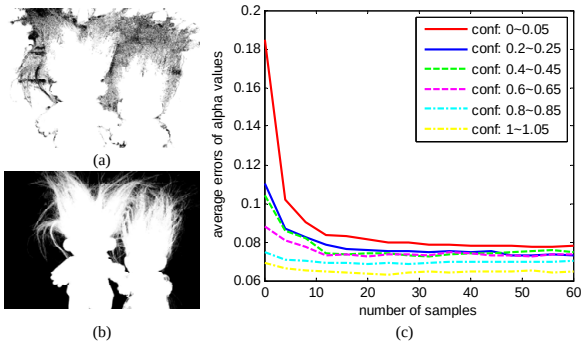


Fig. 4 The confidence image (a), alpha matte (b) and average estimation errors of alpha values versus the number of samples for pixels with different confidence values (c).

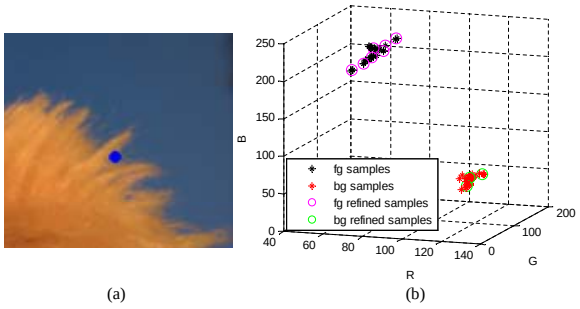


Fig. 5 An example of our adaptive sample clustering strategy. In (b), samples distribution in color space is declared, including un-refined samples (solid points) and refined ones (hollowed-out points).

are similar, estimated alpha values are close to each other which implies that only a few discriminative samples are needed. We employ the K-means clustering to reduce redundant samples. The distance threshold of two clusters is denoted as T_s. As shown in Fig.5, 30 foreground samples and 30 background ones decrease to 7 and 3 respectively. It indicates that only 21 sample-pairs are preserved from 900 ones, where the computational cost decreases significantly.

5. Fine Matting

In this section, we concentrate on fine matting with low computational cost for smoother alpha matte. The matting problem can be described as a MRF model:

$$\min \sum_i (\lambda_i \cdot \mathcal{D}(\alpha_i) + \mathcal{V}(\alpha_i, \boldsymbol{\alpha}_i^N)), \tag{7}$$

where λ_i refers to the balance coefficient between the data term $\mathcal{D}(\alpha_i)$ and smoothness one $\mathcal{V}(\alpha_i, \cdot)$ for the pixel p_i. $\boldsymbol{\alpha}_i^N = \{\alpha_{i_1}, \alpha_{i_2}, ..., \alpha_{i_n}\}$ denotes the pixels in p_i's neighborhood. Our data term and smooth one are declared as,

$$\begin{aligned} \mathcal{D}(\alpha_i) &= \|\alpha_i - \hat{\alpha}_i\| \\ \mathcal{V}(\alpha_i, \alpha_j) &= \|\alpha_i - \sum_j \omega_{ij} \cdot \alpha_j\|, \ p_j \in \mathcal{N}(p_i) \ , \end{aligned} \tag{8}$$

where ω_{ij} evaluates the connectivity of the pixel p_i and p_j.

To avoid the high time and space complexity of the global optimization [13, 14], this paper novelly formulates fine matting into a de-noising problem, where the data term and smoothness one are minimized iteratively and locally.

5.1 Optimization Criterion

Assume the observation error between $\mathbf{I}_i$ and $\hat{\mathbf{I}}_i$ is $\mathbf{n}_i$. According to the linear model in Eq.1, the data term can be re-written as,

$$\mathcal{D}(\alpha_i)=\|\mathbf{I}_i-\hat{\mathbf{I}}_i\|/\|\mathbf{F}_i-\mathbf{B}_i\|=\|\mathbf{n}_i\|/\|\mathbf{F}_i-\mathbf{B}_i\|. \tag{9}$$

Since $\mathbf{F}_i$ and $\mathbf{B}_i$ calculated by color-sampling are regarded as constant, minimizing the data term is equivalent to reducing the observation error $\mathbf{n}_i$. According to Wang [1], the alpha value α_i can be refined by reducing the state error m_i between neighboring ones. The smoothness term is defined by,

$$\sum_j \mathcal{V}(\alpha_i,\alpha_j)=\|\alpha_i-\sum_j \omega_{ij}\cdot\alpha_j\|=\|m_i\|. \tag{10}$$

Hence, we adopt a local de-noising algorithm on neighboring alpha values to minimize the smoothness term. In the meantime, the affinity of the alpha matte is guaranteed. In conclusion, we can convert the objective function in Eq.7 to a de-noising problem as,

$$\begin{aligned}&\min\sum_i(\|\mathbf{n}_i\|+\lambda_i\cdot\|m_i\|)\\ &s.t.\begin{cases}\alpha_i=\sum_j\omega_{ij}\cdot\alpha_j+m_i\\ \mathbf{I}_i-\mathbf{B}_i=(\mathbf{F}_i-\mathbf{B}_i)\cdot\alpha_i+\mathbf{n}_i\end{cases}.\end{aligned} \tag{11}$$

5.2 Iterative Local Optimization

To avoid high time and space complexity of global optimization, we propose to de-noise the alpha matte by minimizing the data term and smoothness one iteratively. Since the optimization problem in Eq.11 can be solved by reducing the error of state and observation equations, this paper employs the standard Kalman filter to work it out. Hence, Eq.11 can be re-described by,

$$\begin{aligned}&\min\sum_i(\|\mathbf{n}_i\|+\lambda_i\cdot\|m_i\|)\\ &s.t.\begin{cases}\alpha_i^k=\sum_j\omega_{ij}\cdot\alpha_j^{k-1}+m_i\\ \mathbf{I}_i-\mathbf{B}_i=(\mathbf{F}_i-\mathbf{B}_i)\cdot\alpha_i^k+\mathbf{n}_i\end{cases}.\end{aligned} \tag{12}$$

Coarse matting has provided us the estimated alpha value, foreground and background components for each pixel, which are the good initial guess of the observation equation. Hence, we minimize the state and observation errors iteratively, until the decline rate of the average alpha value variance is less than the threshold T_f.

1) Minimizing the State Error: Since pixels with more similar color/depth values indicate closer alpha values, we adopt the color and depth similarity to evaluate the connectivity of neighboring ones. The alpha value of the pixel p_i in $(k+1)^{th}$ iteration is estimated by its neighbors,

$$\alpha_i^{k+1|k}=\sum_j\omega_{ij}\alpha_j^{k|k}/\sum_j\omega_{ij}, \tag{13}$$

where $\omega_{ij}=c_j\cdot exp\{-\|\mathbf{I}_i-\mathbf{I}_j\|^2/I_{max}^2\}$ refers to the color similarity between the pixel p_i and p_j, $p_j\in\mathcal{N}(p_i)$ with the window size w_f. Assuming each alpha value is an independent random variable, the variance of the alpha value is denoted by,

$$v_i^{k+1|k}=(\sum_j(\omega_{ij}(\alpha_j^{k|k}-\alpha_i^{k+1|k}))^2)/(\sum_j\omega_{ij})^2+Q_n, \tag{14}$$

where Q_n refers to the variance of the state error.

2) Minimizing the Observation Error: The alpha value of the pixel p_i in $(k+1)^{th}$ iteration will also be updated by the observation equation which is declared as,

$$\alpha_i^{k+1|k+1}=\alpha_i^{k+1|k}+\mathbf{g}_i^{k+1|k^T}(\mathbf{I}_i-\mathbf{B}_i-(\mathbf{F}_i-\mathbf{B}_i)\alpha_i^{k+1|k}). \tag{15}$$

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

$$\begin{aligned}\mathbf{g}_i^{k+1|k^T}&=\frac{v_i^{k+1|k}(\mathbf{F}_i-\mathbf{B}_i)^T}{(\mathbf{F}_i-\mathbf{B}_i)v_i^{k+1|k}(\mathbf{F}_i-\mathbf{B}_i)^T+\mathbf{R}_n},\\ v_i^{k+1|k+1}&=v_i^{k+1|k}[1-\mathbf{g}_i^{k+1|k^T}(\mathbf{F}_i-\mathbf{B}_i)]\end{aligned} \tag{16}$$

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

Table 1 Fixed parameter values in our experiments.

Symbol	Explanation	Value
σ_c	coefficient of confidence values	0.1
T_a	threshold of gradient values	40
N_{min}	minimum number of samples	10
N_{max}	maximum number of samples	30
T_s	threshold of two classes' difference	5
T_f	threshold of the decline rate	0.05
w_f	window size of the smooth filter	5
D_{max}	coefficient of depth values	5
Q_n	the variance of the state error	0.1
$\mathbf{R}_n$	the variance of the observation error	$0.2\cdot\mathbf{I}_{3\times3}$

6. Experiments and Discussions

To evaluate our matting algorithm, we implement and run it on the Intel Core II CPU with 2.0 GHz Dual. Parameter values are fixed as Tab.1. This section comprises 4 parts. *a*) The coarse matting modules are verified. *b*) We evaluate the modules of fine matting. *c*) The overall matting algorithm is compared with other state-of-the-art methods. *d*) We discuss the extension and limitation of our approach. Our algorithm mainly compares with Wang [13], Gastal [15] and He [17] which are representative in local matting. Matting results are tested on the image dataset provided by Rhemann [19], which is regarded as a benchmark for comparison of matting methods. The image dataset includes 27 groups of color images, trimaps with the ground truths.

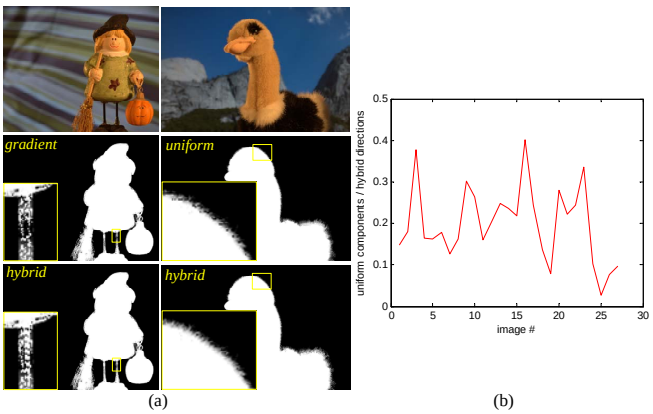


Fig. 6 Visual comparison of the initialization with uniform [15], gradient [17] and hybrid directions (ours) is shown (a). The percentage of uniform components in hybrid directions is also plotted versus the image index (b).

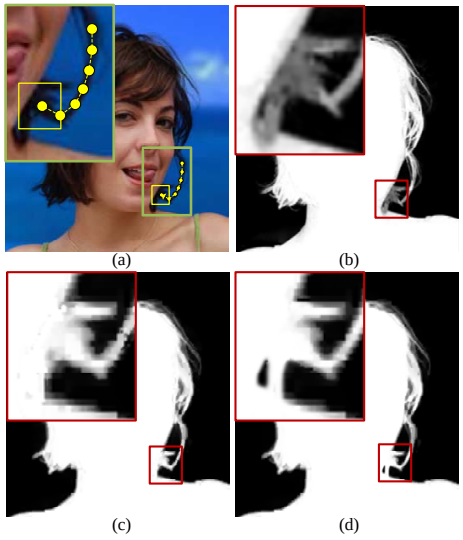


Fig. 7 An example of our sample-pair refinement strategy (d), compared to Gastal [15] (b) and He [17] (c). The yellow route in (a) shows the refinement order.

6.1 Coarse Matting

1) Sample-pair Initialization: As shown in the top row of Fig.6(a), gradient directions of the doll's legs are horizontal (amplitudes are close to 0) and thus imprecise estimations are achieved [17]. In the meantime, non-smooth alpha values are calculated [15], as illustrated in the bottom row of Fig.6(a). Our hybrid directions can balance the effectiveness and discrimination to collect samples well, where more accurate and smoother results are obtained. Fig.6(b) gives the percentage of uniform components in hybrid directions. We can figure out that mostly gradient directions are utilized to gather effective samples. However, uniform directions are also employed to collect discriminative samples, which is essential for the following sample-pair refinement.

2) Sample-pair Refinement: As shown in Fig.7(a), the pixels near the tongue are far away from the foreground hair region. Therefore, imprecise refinement is achieved according to sample-pair refinement [15], as the enlarged region in Fig.7(b). Though He [17] can propagate more sample-pairs of high confidence from pixels far away, the usable foreground component cannot be split to re-construct a sample-pair of higher confidence. Non-smooth matting results are reported in Fig.7(c). However, both high confidence sample-pairs and usable foreground/background components can be utilized in our algorithm. Hence, we can provide more accurate and smoother alpha matte, as the enlarged region in Fig.7(d). The route of our sample-pair refinement is illustrated by yellow points in Fig.7(a).

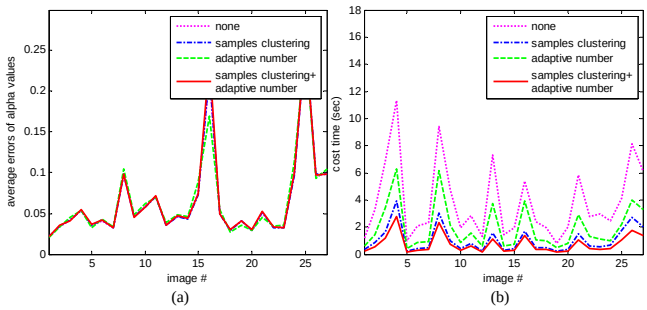


Fig. 8 The average errors of alpha values (a) and the corresponding computational time (b) versus the image index. Different combinations of our speedup strategy are compared.

3) Speedup of Coarse Matting: Fig.8 illustrates the average errors of alpha values and the corresponding computational time versus the image index compared in different combinations of our speedup strategy. We compare our results to not only coarse matting without speedup, but also our method's two variants: *a*) only sample clustering; and *b*) only the adaptive number of samples. Due to the consistency of the image, samples collected from neighbors are of high redundancy, where our sample clustering scheme can speed up significantly. Additionally, the adaptive number of samples helps to reduce sample-pairs of high confidence pixels further. It's more effective specially when there exist a large number of high confidence pixels after initialization. Consequently, our adaptive sample clustering strategy can speed up coarse matting 6 times, while retaining the accuracy of estimations.

4) Evaluation of Coarse Matting: The average errors of alpha values and the corresponding computational time among different coarse matting algorithms are plotted in Fig.9. Wang [13] collected dense samples to estimate alpha values and did not consider the affinity of neighboring pixels. Thus, inaccurate results are reported. In the meantime, if the unknown pixel was far away from foreground/background regions, samples are difficult to collect, causing expensive computational cost. Similar to our algorithm, Gastal [15] achieved coarse matting by means of sample-pair initialization and refinement. Thus, more accurate alpha values were obtained, as shown in Fig.9(a). However, due to the redundancy existing in neighboring samples,

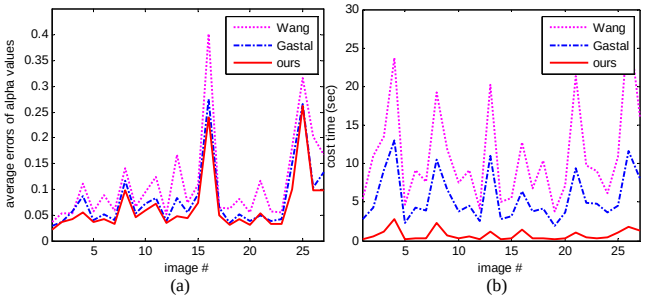


Fig. 9 The average errors of alpha values (a) and computational time (b) versus the image index. Matting results of Wang [13], Gastal [15] and ours are compared.

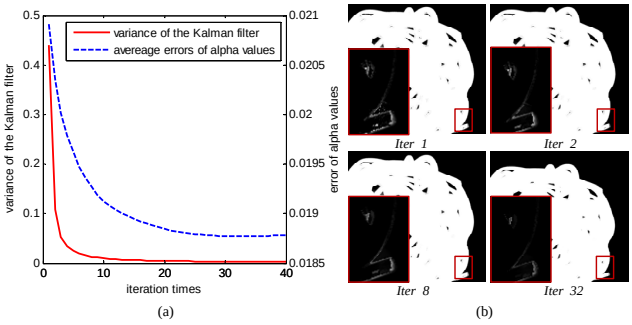


Fig. 10 The variance of our Kalman filter along with the error of alpha values versus the iteration times (a) and several intermediate results of our iterative fine matting (b).

Gastal [15] cannot provide fast estimations.

During sample-pair initialization, our hybrid directions can help to collect effective and discriminative samples. Then we refine sample-pairs considering both high confidence sample-pairs and usable foreground/background components in the neighborhood. Moreover, the decreasing order of the refinement guarantees low confidence pixels are refined after the refinement of more high confidence sample-pairs in the neighborhood. This leads to more accurate estimations of low confidence pixels. To deal with the expensive computational cost for coarse matting, this paper designs an adaptive sample clustering strategy. As illustrated in Fig.9(b), our algorithm is 7 times faster than Gastal [15], and 11 times faster than Wang [13] averagely. To sum up, our algorithm not only achieves more accurate alpha values, but also takes up much lower computational cost.

6.2 Fine Matting

1) Convergence and Effectiveness: To verify the convergence and effectiveness of our fine matting, the variance of our Kalman filter and the average errors of alpha values versus the iteration times are shown in Fig.10(a). The curves indicate that the variance and the corresponding errors decrease simultaneously with the increase of iterations and the iterative process terminates when the iteration time is larger than 20. Our average iterations for matting range from 5 to 20. Fig.10(b) provides several intermediate results of our

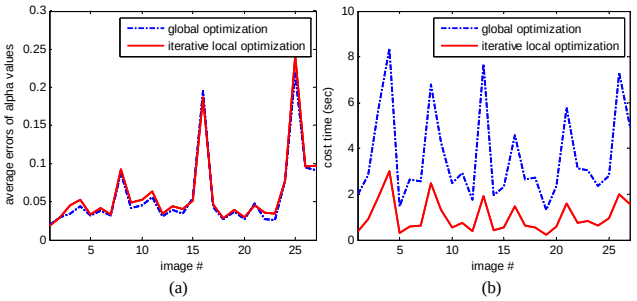


Fig. 11 The average errors of alpha values (a) and the corresponding computational time (b) between global optimization [13, 14] and our iterative local optimization.

fine matting. During the iteration, we reduce the errors of alpha values significantly, as shown in the enlarged regions.

2) Evaluation of Fine Matting: Compared with the global optimization [13, 14], our iterative local optimization costs much lower time and space complexity. It's noted that the initialized alpha mattes are the same for the comparison of different optimization methods. As shown in Fig.11, the average errors of alpha values in our iterative optimization are comparable to the global optimization. In the meantime, our algorithm is 4 times faster than the global optimization [13, 14]. Additionally, for n unknown pixels, the global optimization requires $O(n^2)$ space complexity to construct the matting Laplacian matrix [7]. However, our iterative local optimization minimizes the observation and state errors alternatively, where only $O(n)$ space complexity is needed to store several temporary images.

6.3 Coarse-to-Fine Matting

1) Visual Comparison: Fig.12 visually compares matting results of Wang's [13] (d), Gastal's [15] (e), He's [17] (f) and ours (g). According to the website [20], He's method [17] ranked top in recent local matting techniques.

For the "GT04" image, Wang [13], Gastal [15], He [17] and we achieve clear matting results. Since the background is confusing with the hair of the foreground, Wang's [13] and He's [17] results are imprecise, as shown by the arrows. Gastal's [15] alpha matte is noisy, since the local continuity is not considered. Among all, our algorithm performs better results, as illustrated in the enlarged regions.

For the "GT06" image, as shown in enlarged regions, Wang [13] collects pseudo background samples based on the minimal Euclidean distance, which lead to inaccurate estimations. On the contrary, Gastal's [15], He's [17] and our methods share more sample-pairs for unknown pixels to calculate accurate alpha mattes. Among the three, we achieve the most precise and smoothest results.

For the "GT21" image, the hair of the foreground is confusing with the background, where Wang [13] and He [17] fail to distinguish those pixels well. Gastal's [15] and our approaches reduce the influence of the background. Additionally, we achieve more precise and robust estimations

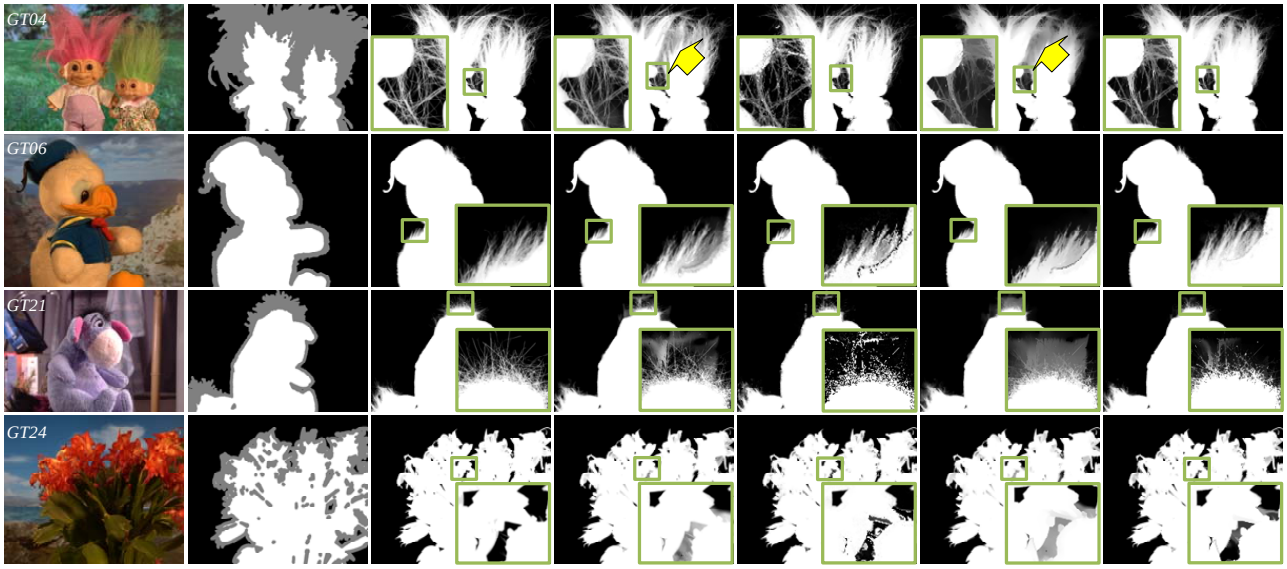


Fig. 12 Visual matting results of Wang's [13] (d), Gastal's [15] (e), He's [17] (f) and ours (g). (a)-(c) refer to the color images, trimaps and ground truths respectively. From the top to bottom rows, images named as "GT04", "GT06", "GT21" and "GT24" are listed respectively.

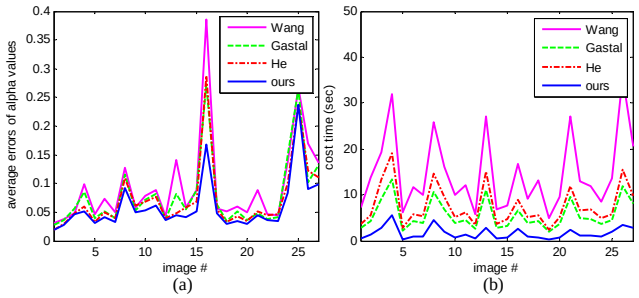


Fig. 13 The average errors of alpha values (a) and computational time (b) versus the image index. Coarse-to-fine matting results of Wang's [13], Gastal's [15], He's [17] and ours are compared.

compared with others, as shown in the enlarged regions.

For the "GT24" image, Wang [13], Gastal [15], He [17] and we can separate the hollowed-out regions. However, Wang [13] and He [17] report inaccurate alpha values, as shown in the enlarged regions. Matting results of Gastal's [15] are noisy, where the affinity of the alpha matte is not guaranteed. Among the above four algorithms, ours reduces the influence of the background and gets the most robust results.

2) Quantitative comparison: The corresponding quantitative comparison between above algorithms [13, 15, 17] is shown in Fig.13. Although Wang [13] applied the global optimization to refine alpha mattes, but the mistaken pixels collected by color-sampling cannot be corrected. Thus, the average errors of alpha values are unbearable. Gastal [15] and He [17] performed a strategy to refine sample-pairs for coarse matting, so that good matting results are reported. As demonstrated in Fig.13(a) our algorithm achieves the most accurate alpha values.

Table 2 The average ranking scores and overall ranks (in brackets) among all algorithms [20] (on the date 2012-09-05).

Algorithm	SAD	MSE
Wang [13]	12.7 (15^{th})	12 (15^{th})
Gastal [15]	9 (11^{th})	10.1 (12^{th})
He [17]	8.8 (10^{th})	8.8 (8^{th})
our algorithm	**6.7** (**4**th)	**7.6** (**6**th)

Further, as declared in Fig.13(b), we cost the shortest computational time. Averagely, our algorithm is 4 times faster than Gastal [15], 5 times faster than He [17] and 8 times faster than Wang [13]. In coarse matting, we propose the adaptive sample clustering strategy for speedup. Consequently, most redundant samples are reduced compared with other color-sampling methods [13, 15, 17]. Moreover, we replace the global optimization of fine matting with our iterative local optimization, which is faster as declared above.

3) Ranks on the benchmark: The website [20] evaluates matting results of top 21 algorithms (on the date 2013-09-05). Average ranking scores in SAD (Sum of Absolute Difference) and MSE (Mean Square Error) evaluation of Wang [13], Gastal [15], He [17] and ours are listed in Tab. 2, where we present the best matting results. Further, among all methods, our algorithm ranks 4 and 6 in SAD and MSE respectively (in brackets). Other matting techniques with better results than ours all require global optimization which confronts expensive time and space complexity, as verified above. In conclusion, our coarse-to-fine matting algorithm outperforms other local matting methods.

6.4 Extension

1) Video Matting: Recent methods [21–23] utilized

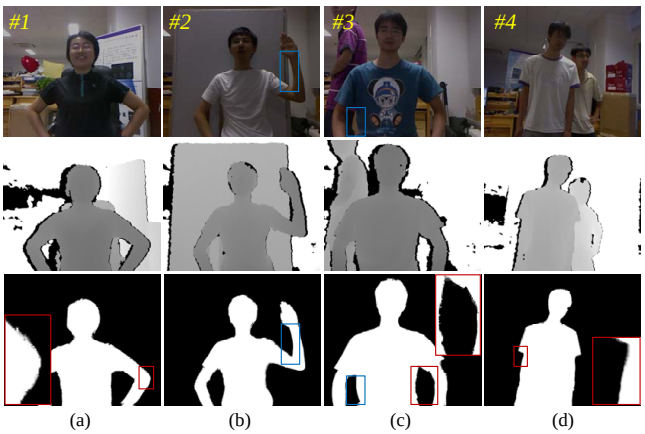


Fig. 14 From the top to bottom rows, the optical image, the depth image and our matting results are illustrated. We take the 22^{th} (#01), 18^{th} (#02), 118^{th} (#03) and 139^{th} (#04) frames as examples.

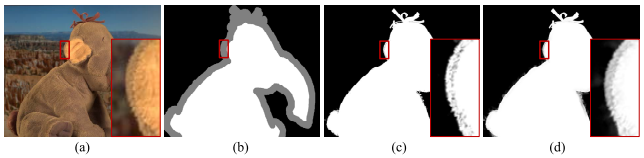


Fig. 15 The color image, trimap and imprecise matting results of ours and Zheng [9]. That is caused by the confusing foreground and background color distributions.

the depth image to bi-layer segment the image, followed by morphology operations on boundaries to generate the trimap automatically. In our experiments we employ the Kinect depth camera [24] to capture images. Then trimaps can be generated [21]. Fig.14 illustrates automatic video matting results of our algorithm. We can figure out that high-accuracy and smooth alpha mattes are achieved by our algorithm.

For frames with the resolution 640×480, we take up $0.17s$ averagely. The computational cost of video matting is much lower than the above image matting. The reason is that the depth camera decreases not only the trivial interaction but also unknown pixels. The number of unknown pixels for image matting ranges from 25,000 to 180,000, but the number for video matting ranges from 10,000 to 20,000. Hence, our algorithm can be extended to a real-time application easily, with some hardware speedups, such as Graphical Processing Unit (GPU) or DSP.

2) Limitation: Our algorithm can provide robust and accurate estimations for alpha values, but it will fail when the foreground and background color distributions are confusing. As declared in Fig.15, inaccurate matting results are achieved since color-sampling methods [13–15, 17] will make mistakes to construct the linear model in Eq.1. Moreover, methods relying on the affinity of the alpha matte [9] cannot estimate accurate results, either, as shown in Fig.15(d). Consequently, distinct foreground and background color distributions are suggested in practical applications of image matting.

7. Conclusion

In this paper, we perform a high-accuracy and quick matting approach. We employ the hybrid (gradient and uniform) directions to shoot rays for foreground/background samples collection. This strategy maximizes the prior knowledge and gathers samples effectively. To achieve precise alpha matte, we refine sample-pairs of unknown pixels with neighboring ones'. Both high confidence sample-pairs and usable foreground/background components are utilized in our method. Moreover, an adaptive sample clustering approach is presented to reduce computational cost of sample-pair selection. Based on sample clustering and the adaptive number of sample-pairs, we speed up coarse matting significantly. Finally, to avoid expensive time and space complexity of global optimization, our algorithm converts it to a de-noising problem. We refine the alpha matte by minimizing the observation and state errors iteratively and locally. Experiments verify the high accuracy and efficiency of our algorithm compared with other state-of-the-art approaches.

Acknowledgments

This work is partially sponsored by NSFC (No. 61271390) and NSFC (No. 61132007).

References

[1] J. Wang and M. Cohen, Image and Video Matting, Now Pub, 2008.

[2] A. Smith and J. Blinn, "Blue screen matting," Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, pp.259–268, ACM, 1996.

[3] Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, "A bayesian approach to digital matting," Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, pp.II–264, IEEE, 2001.

[4] J. Sun, Y. Li, S. Kang, and H. Shum, "Flash matting," ACM Transactions on Graphics (TOG), vol.25, no.3, pp.772–778, 2006.

[5] J. Sun, J. Jia, C. Tang, and H. Shum, "Poisson matting," ACM Transactions on Graphics (ToG), vol.23, no.3, pp.315–321, 2004.

[6] J. Wang and M. Cohen, "An iterative optimization approach for unified image segmentation and matting," Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, pp.936–943, IEEE, 2005.

[7] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, pp.61–68, IEEE, 2006.

[8] X. Bai and G. Sapiro, "A geodesic framework for fast interactive image and video segmentation and matting," Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pp.1–8, IEEE, 2007.

[9] Y. Zheng and C. Kambhamettu, "Learning based digital matting," Computer Vision, 2009 IEEE 12th International Conference on, pp.889–896, IEEE, 2009.

[10] Q. Chen, D. Li, and C. Tang, "Knn matting," Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp.869–876, IEEE, 2012.

[11] P. Lee and Y. Wu, "Nonlocal matting," Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp.2193–2200, IEEE, 2011.

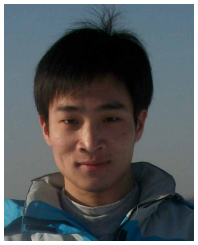
[12] Y. Guan, W. Chen, X. Liang, Z. Ding, and Q. Peng, "Easy matting-a stroke based approach for continuous image matting," Computer Graphics Forum, pp.567–576, Wiley Online Library, 2006.
[13] J. Wang and M. Cohen, "Optimized color sampling for robust matting," Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pp.1–8, IEEE, 2007.
[14] C. Rhemann, C. Rother, and M. Gelautz, "Improving color modeling for alpha matting," British Machine Vision Conference, pp.1155–1164, 2008.
[15] E. Gastal and M. Oliveira, "Shared sampling for real-time alpha matting," Computer Graphics Forum, pp.575–584, Wiley Online Library, 2010.
[16] K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun, "A global sampling method for alpha matting," Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp.2049–2056, IEEE, 2011.
[17] B. He, G. Wang, Z. Ruan, X. Yin, X. Pei, and X. Lin, "Local matting based on sample-pair propagation and iterative refinement," Image Processing (ICIP), 2012 IEEE Conference on, IEEE, 2012.
[18] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann, "Random walks for interactive alpha-matting," Proceedings of VIIP, pp.423–429, 2005.
[19] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott, "A perceptually motivated online benchmark for image matting," Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp.1826–1833, IEEE, 2009.
[20] "Alpha matting benchmark." http://www.alphamatting.com, 2009.
[21] J. Zhu, M. Liao, R. Yang, and Z. Pan, "Joint depth and alpha matte optimization via fusion of stereo and time-of-flight sensor," Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp.453–460, IEEE, 2009.
[22] J. Cho, R. Ziegler, M. Gross, and K. Lee, "Improving alpha matte with depth information," IEICE Electronics Express, vol.6, no.22, pp.1602–1607, 2009.
[23] L. Wang, M. Gong, C. Zhang, R. Yang, C. Zhang, and Y. Yang, "Automatic real-time video matting using time-of-flight camera and multichannel poisson equations," International journal of computer vision, vol.97, no.1, pp.104–121, 2012.
[24] "Microsoft kinect for x-box 360." http://www.xbox.com/en-US/kinect, 2010.

Bei He was born in Huaining, Anhui, in 1987. He received the B.S. degree from the department of Electronics Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2008. He is currently working toward the Ph.D. degree in the department of Electronics Engineering, Tsinghua University, Beijing, China. His research interests include the applications of image processing and pattern recognition in image/video matting, registration and mosaicing.

Guijin Wang was born in 1976. He received the B.S. and Ph.D. degree (with honor) from the department of Electronics Engineering, Tsinghua University, China in 1998, 2003 respectively, all in Signal and Information Processing. From 2003 to 2006, he has been with Sony Information Technologies Laboratories as a researcher. From Oct., 2006, he has been with the department of Electronics Engineering, Tsinghua University, China as an associate professor. He has published over 50 International journal and conference papers, hold several patents. He is the session chair of IEEE CCNC'06, the reviewers for many international journals and conferences. His research interests are focused on wireless multimedia, image and video processing, depth imaging, pose recognition, intelligent surveillance, industry inspection, object detection and tracking, online learning, etc.

Chenbo Shi was born in Qidong, Jiangsu, in 1984. He received the B.S., M.S. and Ph.D. degree from the department of Electronics Engineering, Tsinghua University, Beijing, China, in 2005, 2008 and 2012 respectively. He is currently working toward the Post D. degree in the department of Electronics Engineering, Tsinghua University, Beijing, China. His research interests include image/video registration, local features and stereo matching.

Xuanwu Yin was born in Changchun, Jilin, in 1987. He received the B.S. degree from the department of Electronics Engineering, Tsinghua University, Beijing, China, in 2011. He is currently working toward the Ph.D. degree in the department of Electronics Engineering, Tsinghua University, Beijing, China. His research interests include the applications of image processing and pattern recognition in image/video matting, registration and 3D reconstruction.



Bo Liu was born in Shaodong, Hunan, in 1989. He received the B.S. degree from the department of Electronics Engineering, Tsinghua University, Beijing, China, in 2011. He is currently working toward the M.S. degree in the department of Electronics Engineering, Tsinghua University, Beijing, China. His research interests include image/video registration, mosaicing, super-resolution and industrial inspection.

Xinggang Lin received B.S. degree in Electronics Engineering, Tsinghua University, China in 1970; M.S. degree in 1982 and Ph.D. degrees in 1986, both in information science, Kyoto University, Japan. He joined the Department of Electronics Engineering at Tsinghua University in 1986 where he has been a full professor since 1990. He received "Great Contribution Award" from Ministry of Science and Technology of China, and "Promotion Awards of Science and Technology" from Beijing Municipality. He was a General Co-chair of the second IEEE Pacific-Rim Conference on Multimedia, an associate editor of IEEE T. on CSVT, and a technical/organizing committee member of many international conferences. He is a fellow of China Institute of Communications, and he published over 140 referred conference and journal papers in diversified research fields.