Local and Nonlocal Color Line Models for Image Matting

Byoung-Kwang KIM*, Meiguang JIN††, Nonmembers, and Woo-Jin SONG†, Member

SUMMARY In this paper, we propose a new matting algorithm using local and nonlocal neighbors. We assume that K nearest neighbors satisfy the color line model that RGB distribution of the neighbors is roughly linear and combine this assumption with the local color line model that RGB distribution of local neighbors is roughly linear. Our assumptions are appropriate for various regions such as those that are smooth, contain holes or have complex color. Experimental results show that the proposed method outperforms previous propagation-based matting methods. Further, it is competitive with sampling-based matting methods that require complex sampling or learning methods.

Key words: K nearest neighbors, color line model, smoothness, propagation-based matting

1. Introduction

Matting refers to the accurate foreground extraction from images and videos along with the opacity estimation for each pixel. It is widely used in image editing and film production. For the ith pixel, in general, a color value $I_i$ can be modeled as a convex combination of a foreground $F_i$ and background $B_i$ using alpha $\alpha_i$, the foreground opacity [1]:

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i,$$

where $\alpha_i$ has the range [0,1]. The goal of image matting is to estimate the alpha values of the pixels. Matting is a severely underconstrained problem since all variables on the right-hand side of (1) are unknown. Therefore, user interaction, such as a trimap, which consists of three parts: known foreground, known background, and unknown, is often required.

Previous matting algorithms can be divided into propagation-based and sampling-based approaches. The propagation-based approaches define affinities between neighbor pixels [2]–[6]. According to assumptions on the image statistics, they estimate and propagate alpha values from the known foreground and the known background to unknown alpha values. Closed-Form (CF) matting [2] assumes that, within a small window, each of $F$ and $B$ is distributed linearly in the RGB color space, which is called the local color line model, and derives affinities between local neighbors. Recent propagation-based approaches focus on the affinities of nonlocal neighbors. Nonlocal matting [5] and K nearest neighbors (KNN) matting [6] assume that pixels sharing same appearance should be expected to share the same alpha values.

Sampling-based approaches require foreground and background sample pairs [7]–[10]. They assume that the true foreground and background colors of an unknown pixel exist in the image and search for the most appropriate sample pairs among the known foreground and background pixels. Thus, they are dependent on samples and require a lot of information about the foreground and background. Furthermore, the computational complexity is high because of the sample pair comparisons. The results of sampling-based approaches alone may have roughness and discontinuity in smooth regions. Thus, sampling-based approaches usually adopt a post-processing method, combining the results obtained by sample pairs with the matting Laplacian matrix of [2] that can be treated as a smoothness term [9].

In this paper, we propose a new propagation-based matting method using local and nonlocal neighbors. We assume that the KNN of a pixel in feature space satisfy the color line model; we refer this assumption as the nearest neighbors (NN) color line model. A short earlier version of this work appeared in [11]. This earlier algorithm performed matting using only the NN color line model; we refer to it as NN smoothness matting. However, in this work, we combine this method with the local color line model [2]. Since the proposed algorithm accurately represents correlations between local neighbors as well as nonlocal neighbors, it can preserve local smoothness for smooth regions, as well as, propagate foreground and background information correctly to hole and complex color regions. Experimental results show that the proposed algorithm outperforms previous propagation-based matting methods and is competitive with previous sampling-based matting methods that require complex learning or sampling methods.

2. Related Work

The proposed algorithm is based on local and nonlocal propagation methods. In particular, we discuss CF matting [2] and KNN matting [6] which are closely related to our work.

2.1 Closed-Form Matting

CF matting [2] assumes that, within a small window, each
of $F$ and $B$ is distributed linearly in the RGB color space and can be modeled as a linear mixture of two colors. This assumption, called the local color line model, is very appropriate for smooth regions such as fur regions (as in the first example of Fig. 3(d)). However, since it is based on local neighbors, foreground or background information cannot be propagated well in hole regions, that is, unknown regions that are bounded by constraints of one type only (the second example of Fig. 3(d)). Furthermore, it is not appropriate for complex color regions, where color variations are complex, because the color line model is too simple to explain complex color variations (the third example of Fig. 3(d)).

2.2 KNN Matting

Based on the assumption of nonlocal principle, KNN matting [6] assumes that pixels sharing the same appearance should be expected to share the same alpha value. Since these methods are not limited to local neighbors, foreground and background information can be propagated to hole and complex color regions so obtain fine matting results for these regions (the second and third examples of Fig. 3(f)). However, they cannot obtain sufficiently smooth matting results in the fur regions shown in the first example of Fig. 3(f), because they have no smoothness assumption.

3. Nearest Neighbors Color Line Model

As mentioned in the previous section, CF matting [2] can obtain good matting results for smooth regions. It cannot, however, obtain satisfactory results for hole and complex color regions. On the other hand, KNN matting [6] can obtain good matting results for hole regions and complex color regions but cannot perform well for smooth regions. Inspired by CF and KNN matting, we propose a novel algorithm that can obtain high quality matting results for smooth regions, hole regions, and complex color regions. In the proposed algorithm, we assume that the KNN of a pixel in feature space satisfy the color line model as well as the local smoothness assumption that states that local neighbors in a spatial domain satisfy the color line model. In this section, we describe the NN color line model and the cost function from the model.

3.1 Feature Vector

First, we search KNN. For each pixel, we construct feature vectors defined by RGB color and location information as for conventional KNN matting [6]: $X(i) = (I^R, I^G, I^B, x, y)$, where all color and location variables are normalized. Then, we search for KNN in the feature space using the Euclidean difference $||X(i) - X(j)||$. To implement this process easily, we used FLANN [12], a very efficient KNN implementation. Figure 1(b) shows the nonlocal neighbors computed using FLANN.

3.2 NN Color Line Model Derivation

We assume that the KNN in the feature space satisfy the color line model. Figure 1 demonstrates that this assumption can be satisfied for smooth, hole, and complex color regions. Figures 1(c), (d) are the corresponding foreground and background distributions of KNN at a pixel, respectively. Here the foregrounds and backgrounds are reconstructed as in CF matting. They show that the distributions are linearly distributed. Therefore, the color line model can be applied to KNN in the feature space. However, the NN color line model does not mean the feature vectors are linearly distributed. Actually, only the color components of the KNN are linearly distributed, hence we make the NN color line model. Since the foreground and background of the KNN are assumed to satisfy the color line model, they can be represented by the convex combination of two colors: $∀ j \in N_i, F_j = \beta_{j}^F F_1 + (1 - \beta_{j}^F) F_2$, and $B_j = \beta_{j}^B B_1 + (1 - \beta_{j}^B) B_2$, where $N_i$ is the neighbor vector composed of the KNN of pixel $i$. $F_1$, $F_2$, $B_1$, and $B_2$ are constant colors over $N_i$. Therefore, the color of a pixel $∀ j \in N_i$ can be denoted as

$$I_j = \alpha_j B_j^F F_1 + (1 - \beta_j^F) F_2 + (1 - \alpha_j)(\beta_j^B B_1 + (1 - \beta_j^B) B_2), ∀ j \in N_i. \quad (2)$$

As in CF matting, we can obtain a 4D linear model from (2)

$$\alpha_j = a_j^R I_j^R + a_j^G I_j^G + a_j^B I_j^B + b_j, ∀ j \in N_i. \quad (3)$$

3.3 Cost Function from NN Color Line Model

Using the 4D linear model in (3), the linear coefficients can be estimated by minimizing the cost function

$$J_n = \sum_{i \in I} \left( \sum_{j \in N_i} (\alpha_j - \sum_c a_j^c I_j^c - b_j)^2 + \epsilon \sum_c (d_j^c)^2 \right). \quad (4)$$

where $I$ is the entire image, $C$ is the color channel and $\epsilon$ is
a small constant for a regularization term on $\alpha_c^2$. The cost function is quadratic in $\alpha$, $\alpha_c^2$, and $b_i$ with $5N$ unknowns for an image with $N$ pixels. However, $\alpha_c^2$ and $b_i$ can be eliminated from (4) using Theorem 1 in [2], yielding the cost function dependent only on $\alpha$ with $N$ unknowns as

$$J_n = \alpha^T L_n \alpha.$$ (5)

Here, $\alpha$ is an $N \times 1$ vector, and $L_n$ is an $N \times N$ Laplacian matrix, whose $(j,k)$th element is

$$\sum_{i,j,k\in N} \left( \delta_{jk} - \frac{1}{|N|} \right) \left( 1 + \frac{(I_j - \mu_i)^T (S_i + \frac{\epsilon}{|N|} I_3)^{-1} (I_k - \mu_i) }{\epsilon} \right),$$ (6)

where $N$ is the number of pixels in an image, $S_i$ is a $3 \times 3$ covariance matrix, $\mu_i$ is the $3 \times 1$ mean vector of the intensities in the neighbor vector $N_i$, $I_3$ is the $3 \times 3$ identity matrix.

The NN color line model is effective. First of all, it is not limited to color variations. Because color variations differ for different images, previous matting methods do not cover these variations. Based on the nonlocal principle, our model does not depend on color variations and can take into account these variations. Second, the NN color line model can reflect variations of the foreground and background. Although KNN matting is also based on the nonlocal principle, it assumes that pixels sharing the same appearance should be expected to share the same alpha value, which is not always appropriate, especially for smooth regions. From (1), we know that alpha mattes are closely related to the foreground and background, but KNN matting bases its assumption directly on alpha values without considering the foreground or background. The KNN of a pixel in a fur region are shown in Fig. 1(b). We found that the ground truth alpha values of the KNN are not very similar because of the foreground and background variations that violate the assumption in KNN matting. The proposed algorithm, based on the color line model, can reflect these variations in foreground and background and estimate alpha mattes more accurately.

4. Combining Nearest Neighbors Color Line Model with Local Color Line Model

Although the NN color line model explains the correlation between nonlocal neighbors better than KNN matting, the results from this model are not as smooth as the results of CF matting for smooth regions (the first example of Fig. 3). Therefore, we apply the local color line model as well as the NN color line model to obtain smooth matting results for smooth regions. For hole regions, if we apply only the local color line model, the results tend to be oversmooth because it biases the mattes to be locally constant [4]. For example, for hole regions bounded by only known foreground regions in a trimap, known background information cannot be propagated locally and then the estimated alpha mattes of the hole regions become oversmooth. Since our method can propagate information nonlocally as well as locally, background information can be propagated to the hole regions.

Therefore, the alpha mattes are not biased and can be estimated more accurately (the second example of Fig. 3). For complex regions, if the local neighbor size is large, the local color line model may not be appropriate. However, the local window size is small ($3 \times 3$ in CF matting and our algorithm) and therefore the local color line model does not considerably affect the matting results (the third example of Fig. 3). These are similar to the sampling-based approaches [7]–[9] that usually adopt a post-processing method by adding the matting Laplacian matrix of [2] which can be treated as a smoothness term to get smooth results for smooth regions without degrading hole and complex color regions. The cost function combining $J_n$ from the NN color line model with the cost function $J_l$ from the local color line model [2] is given by

$$J = \mu J_n + (1 - \mu) J_l + \lambda (\alpha^T - b_5^T) D_5 (\alpha - b_5)$$

$$= \mu \alpha^T L_n \alpha + (1 - \mu) \alpha^T L_l \alpha + \lambda (\alpha^T - b_5^T) D_5 (\alpha - b_5),$$ (7)

where $L_l$ is Laplacian matrix from local color line model [2], $\lambda$ is a large number, $D_5$ is a diagonal matrix whose diagonal elements are 1 for constrained pixels and 0 for all other pixels, and $b_5$ is the vector containing the specified alpha values for the constrained pixels and zero for all other pixels. $\mu$ is a constant between 0 and 1. Then, we minimize the
combined cost function by differentiating (7) and setting the derivatives to zero. The final alpha is
\[
\alpha = (\mu L_m + (1 - \mu)L_l + AD_\alpha)^{-1}\lambda b_s.
\] (8)

\(\mu\) is a user parameter which balances between the local and NN color line models. If \(\mu\) is not a extreme value such as 0 or 1, experimental results are not sensitive to \(\mu\). Figure 2 shows a qualitative comparison under different values of \(\mu\). We set \(\mu\) to be 0.5 in our experiments. Table 1 shows the summary of proposed algorithm.

5. Experimental Results

In this section, we compare the performance of the proposed algorithm to other state-of-the-art matting methods. All experimental images, including the original images and trimaps, were obtained from the benchmark dataset in [13]. They consisted of 27 training images with ground truth and 8 test images without ground truth. We set \(K = 9\), \(\epsilon = 10^{-7}\), \(\mu = 0.5\), and \(\lambda = 100\).

Figure 3 shows the matting results of the training images. For smooth regions (the first example in Fig. 3), the KNN matting result looks like a hard segmentation because it has no smoothness assumption. However, the result from the proposed algorithm is much smoother than KNN matting and closer to the ground truth. Although NN smoothness matting, which is based on the NN color line model only, gives a smoother result than KNN matting, the proposed algorithm has the local color line model as well as the NN color line model that can estimate alpha mattes accurately in smooth regions. For hole regions, since CF matting uses only local information, if there is no background information in the hole, an estimated alpha is biased towards the foreground. However, using nonlocal neighbors, the proposed algorithm can propagate background information to the hole and thus estimate alpha mattes very well (the second example in Fig. 3). For the complex regions, the proposed algorithm shows high quality performance because of its nonlocal principle (the third example in Fig. 3).

Matting results of the test images are shown in Fig. 4. We compared the proposed algorithm with SVR matting [9], a kind of sampling-based matting that was ranked highly in

<table>
<thead>
<tr>
<th>Table 2</th>
<th>MSE of the matting results.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>2.18 x 10^{-2}</td>
</tr>
<tr>
<td>SAD</td>
<td>5.32 x 10^3</td>
</tr>
</tbody>
</table>
As shown in the first example of Fig. 4, the proposed algorithm had a high performance, although the foreground object’s hair overlapped with the background structure. The SVR matting is dependent on the trimap since it is based on sampling and a learning algorithm. It can be vulnerable to errors if sample candidates do not contain the true foreground and background (the second example in Fig. 4). For textured regions, the proposed algorithm preserves smoothness and estimates more accurately than the other methods (the third example in Fig. 4).

For quantitative comparison, we computed the mean squared error (MSE): \( \frac{1}{N_{un}} \sum (\alpha_i - \hat{\alpha}_i)^2 \), and the sum of absolute difference (SAD): \( \sum |\alpha_i - \hat{\alpha}_i| \), where \( \alpha_i \) and \( \hat{\alpha}_i \) are the estimated and ground truth alpha mattes of pixel \( i \), respectively, and \( N_{un} \) is the number of unknown pixels in the trimap. Table 2 shows that the proposed algorithm, which combines the NN color line model and the local color line model, has outstanding performance when compared to the other methods. In addition, we evaluated our performance for the test images in [13]. Table 3 shows the rank of MSE for many matting methods. The proposed algorithm has the best performance among the propagation-based matting methods and is competitive with the sampling-based matting methods that require complex sampling processes. For some images, in fact, the proposed algorithm has the best results among all matting methods in [13].

Table 3  Rank of MSE as well as the normalized score for test images on the benchmark dataset [13].

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Rank of MSE</th>
<th>Normalized Score(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>overall user small</td>
<td></td>
</tr>
<tr>
<td>SRLO matting [10]</td>
<td>16.3 6.3 15.9 15</td>
<td>14 3.9 11 13</td>
</tr>
<tr>
<td>Learning based matting [3]</td>
<td>16.6 15 15.8 19.1</td>
<td>22 8 24 24</td>
</tr>
<tr>
<td>CF matting [2]</td>
<td>16.6 15 15.8 19.1</td>
<td>22 8 24 24</td>
</tr>
<tr>
<td>Proposed algorithm</td>
<td>11.4 13.4 10.6 10.1</td>
<td>1 1 1 3 3 69.0</td>
</tr>
<tr>
<td>SVR matting [9]</td>
<td>8.2 10.6 7 7</td>
<td>24 28 4 15 9 73.0</td>
</tr>
<tr>
<td>Shared matting [8]</td>
<td>11.8 10.8 13.9 10.6</td>
<td>4 15 11 6 11 58.2</td>
</tr>
<tr>
<td>KNN matting [6]</td>
<td>12.2 14 12.1 10.5</td>
<td>19 6 5 12 22 63.6</td>
</tr>
</tbody>
</table>

Acknowledgments

This work was supported in part by the NRF grant funded by MEST (2012R1A2A2A01011112) and in part by the MSIP under the CITRC support program (NIPA-2014-H0401-14-1001) supervised by the NIPA.

References


