# Image Matting with KL-Divergence Based Sparse Sampling

Levent Karacan Aykut Erdem Erkut Erdem
Department of Computer Engineering, Hacettepe University
Beytepe, Ankara, TURKEY, TR-06800
{karacan,aykut,erkut}@cs.hacettepe.edu.tr

## Abstract

*Previous sampling-based image matting methods typically rely on certain heuristics in collecting representative samples from known regions, and thus their performance deteriorates if the underlying assumptions are not satisfied. To alleviate this, in this paper we take an entirely new approach and formulate sampling as a sparse subset selection problem where we propose to pick a small set of candidate samples that best explains the unknown pixels. Moreover, we describe a new distance measure for comparing two samples which is based on KL-divergence between the distributions of features extracted in the vicinity of the samples. Using a standard benchmark dataset for image matting, we demonstrate that our approach provides more accurate results compared with the state-of-the-art methods.*

## 1. Introduction

Accurately estimating foreground and background layers of an image plays an important role for many image and video editing applications. In the computer vision literature, this problem is known as image matting or alpha matting, and mathematically, refers to the problem of decomposing a given image $I$ into two layers, the foreground $F$ and the background $B$, which is defined in accordance with the following linear image composition equation:

$$I = \alpha F + (1 - \alpha)B \tag{1}$$

where $\alpha$ represents the unknown alpha matte which defines the true opacity of each pixel and whose values lies in $[0, 1]$ with $\alpha = 1$ denoting a foreground pixel and $\alpha = 0$ indicating a background pixel. This is a highly ill-posed problem since for each pixel we have only three inputs but seven unknowns ($\alpha$ and the RGB values of $F$ and $B$). The general approach to resolve this issue is to consider a kind of prior knowledge about the foreground and background in form of user scribbles or a trimap to simplify the problem and use the spatial and photometric relations between these known pixels and the unknown ones.

Image matting methods can be mainly categorized into two groups: propagation-based methods [23, 10, 16, 15, 3, 22, 11] and sampling-based methods [6, 27, 9, 12, 20, 21, 25, 13]. The first group defines an affinity matrix representing the similarity between pixels and propagate the alpha values of known pixels to the unknown ones. These approaches mostly differ from each other in their propagation strategies or affinity definitions. The latter group, on the other hand, collects color samples from known foreground and background regions to represent the corresponding color distributions and determine the alpha value of an unknown pixel according to its closeness to these distributions. Early examples of sampling-based matting methods [6, 27] fit parametric models to color distributions of foreground and background regions. Difficulties arise, however, when an image contains highly textured areas. Thus, virtually all recent sampling-based approaches [9, 12, 20, 21, 25, 13] consider a non-parametric setting and employ a particular selection criteria to collect a subset of known $F$ and $B$ samples.Then, for each unknown pixel, they search for the best $(F, B)$ pair within the representative samples, and once the best pair is found, the final alpha matte is computed as

$$\hat{\alpha} = \frac{(I - B).(F - B)}{\|F - B\|^2} . \tag{2}$$

The recent sampling-based approaches mentioned above also apply local smoothing as a post-processing step to further improve the quality of the estimated alpha matte. Apart from the two main types of approaches, there are also some hybrid methods which consider a combination of propagation and sampling based formulations [4], or some supervised machine learning based methods which learn proper matting functions from a training set of examples [29]. For a more comprehensive up-to-date survey of image matting methods, we refer the reader to [30, 26].

The matting approach we present in this paper belongs to the group of sampling-based methods that rely on a non-parametric formulation. As will be discussed in more detail in the next section, these methods typically exploit different strategies to gather the representative foreground and back-

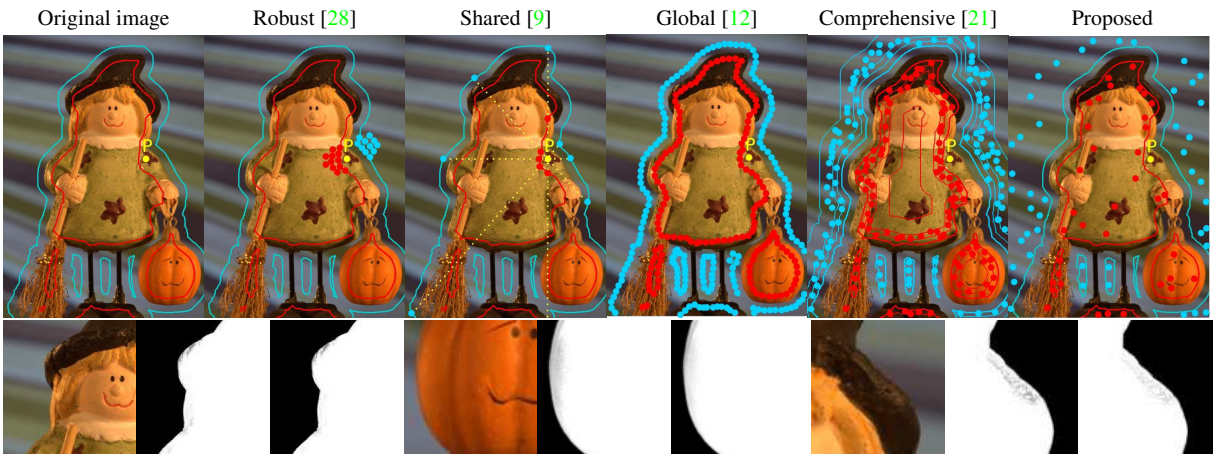



Figure 1. Non-parametric sampling-based matting approaches. Top row: An input image and the representative samples gathered by the Robust [28], Shared [9], Global [12], Comprehensive [21], and the proposed Sparse Sampling based matting methods. The unknown pixel, the foreground and background samples are shown in yellow, red and blue colors, respectively. Bottom row: Comparison of the estimated alpha mattes by the suggested approach and the state-of-the-art Comprehesive Sampling matting method [21].

ground samples. Our observation is that all these strategies lack a strong theoretical basis, *i.e*. they require certain assumptions to hold to capture the true foreground and background colors, and moreover, they fail to adequately utilize the relationship between known and unknown regions. In contrast, our approach offers a more principled way to sampling by casting it as a sparse subset selection problem [7, 8], in which the resulting samples refers to a small subset of known foreground and background pixels that best explains the unknown pixels. In particular, sampling is formulated as a row-sparsity regularized trace minimization problem which solely depends on pairwise dissimilarities between known and unknown pixels, and for that, we propose a new KL-divergence based contextual measure as an efficient alternative to chromatic and spatial distances.

**Previous work.** Sampling-based image matting models mainly differ from each other in (i) how it collects the representative foreground and background samples, and (ii) how it selects the best $(F, B)$ pair for an unknown pixel.

An early example is the Mishima's Blue-screen matting method [17] in which the image of a foreground object is captured in front of a monochrome background. This setup allows efficient estimation of foreground and background distributions via clustering, and then alpha values of unknown pixels are estimated by considering their proximity to the extracted clusters. Another early work, the Knockout system [2], estimates true color values of the foreground and background layers of an unknown pixel by a weighted sum of nearby known pixels with the weights proportional to their spatial distances to the unknown pixel.

Robust matting [28], for an unknown pixel, collects samples from the known nearby foreground and background pixels. Among those samples, it then selects the pair that best fits the linear compositing equation defined in Eq. (1). As the selection is carried by taking into account the color distortion, it provides more robust results than the Knockout system. However, since sampling depends only on the spatial closeness to the unknown pixels, as shown in Fig. 1, the true samples might be missing in the candidate set, decreasing the matting quality. In [18], it has been shown that using geodesic distances improves the results of this model to a certain extent.

Shared matting [9] gathers representative samples from the trimap boundary, assuming that, for an unknown pixel, its true foreground and background color can be found at the closest known region boundaries. These pixels are defined as the boundary pixels that lie along the rays which are originated from the unknown pixel and that partition the image plane into disjoint parts of equal planar angles. Then, the best pair among those are used to estimate its alpha value w.r.t. an objective function that depends on spatial and photometric affinity. It falls short, however, when the rays do not reach the true samples. Weighted color and texture (WCT) sampling [20] and its comprehensive version (CWCT) extend Shared matting by combining the local sampling strategy in [9] with a global one that depends on a clustering-based probabilistic model. Moreover, it employs a texture compatibility measure in addition to the color distortion measure to prevent selecting overlapping samples.

Global sampling [12] also collects samples from the trimap boundaries but to avoid the problem of missing true samples, instead of emanating rays from unknown pixels, as in Shared matting, it considers all known boundary samples as a global candidate set. To handle the large number of samples, it employs a simple objective function and an efficient random search algorithm in finding the best sample pair. However, as shown in Fig. 1, the true colors might still be missed in the resulting sample set if they do not lie

along the trimap boundaries.

Comprehensive sampling matting [21] follows a global strategy and divides the known and unknown regions into a number of segments so that the segment over which the samples are gathered is decided according to the distance of a given unknown pixel to the extracted foreground and background segments. Sample colors are constructed as the means of the color clusters that are obtained via a two-level hierarchical clustering modeled by a parametric Gaussian mixture model. This approach gives better results than the previous non-parametric sampling based approaches. However, there is still a possibility of missing true samples since the sampling strategy depends on spatial closeness. As demonstrated in Fig. 1, the true color samples might be very far away from the unknown pixel.

Sparse coded matting [13] formulates image matting as a sparse coding problem. It computes alpha values from a bunch of sample pairs within a sparse coding framework instead of finding only the best but single pair of foreground and background $(F, B)$ pair. These samples forming the dictionary atoms are collected from the mean color of the superpixels that lie along the boundaries of the trimaps. Thus, it might also suffer from the missing true samples problem. To prevent overlapping color distributions of foreground and background, it adaptively controls the dictionary size according to a confidence value that depends on probabilistic segmentation.

**Our contributions.** As described above, all existing sampling-based image matting methods rely upon different assumptions regarding the selection policy of background and foreground samples. The justification of these assumptions are mostly valid. But still, they are heuristic methods and they all lack a theoretical ground to explain the relationship between known and unknown pixels. As a step towards improving those methods, in this paper we present a new approach for image matting. As shown Fig. 1, the proposed method achieves a more effective sampling and provides considerably better alpha mattes. To conclude the introduction, the main contributions of this paper can be summarized as follows:

(1) To overcome the limitations of the previous works, we develop a well-founded sampling strategy, which rely on a recently proposed sparse subset selection technique [7], to select a small set of foreground and background samples that best explain the unknown pixels.

(2) We design a new measure of distance between two samples based on KL-divergence between the distributions of the features extracted in the vicinity of the samples. This measure is used in both selecting the representative samples and finding the best $(F, B)$ pair for an unknown pixel. As shown in Fig. 1, when combined with our sampling strategy,

(3) We provide compelling qualitative and quantitative results on a benchmark dataset of images [19] that demonstrate substantial improvements in the estimated alpha mattes upon current state-of-the-art methods.

## 2. Proposed Approach

In this study, we build upon a recent work by Elhamifar *et al.* [7], and address the sampling process in image matting as a sparse subset selection problem. In particular, we find a few representative pixels for the known foreground and background regions solely based on pairwise distances between the known and unknown pixels. As in other sampling-based approaches, in our formulation, the distance measure used in comparing two samples is of great importance since it directly affects the quality of selected samples. As we mentioned earlier, another contribution of this study is a new distance which is based on KL-divergence between feature distributions. In the following, we begin with the definition of our distance measure, and then discuss the details of the proposed algorithm. The steps of the algorithm involves collecting foreground and background color samples from known pixels via sparse subset selection, then we define an objective function to find the best $(F, B)$ pair for an unknown pixel according to linear composition equation.

### 2.1. Measure of Distance Between Two Samples

Sampling-based approaches generally consider very simple measures which depend on chromatic and/or spatial similarities [28, 9, 12, 21]. The only exceptions are [20, 25], which also employ some texture similarity measures. Unlike those measures, here, we consider a statistical data representation and propose to use an information-theoretic approach. In particular, our measure depends on a parametric version of the Kullback-Leibler (KL) Divergence [14], a well-known non-symmetric measure of the difference between two probability distributions in information theory, which we describe below. We note that KL-Divergence was used in a different way for video matting previously in [5].

Given an input image, we extract a 9-dimensional feature vector $\phi$ for each pixel as follows:

$$\phi(x, y) = \left[ \begin{array}{ccccccccc} x & y & r & g & b & |I_x| & |I_y| & |I_{xx}| & |I_{yy}| \end{array} \right]^\top \quad (3)$$

with $(x, y)$ denoting the pixel location, $I = [r\ g\ b]$ representing the pixel values of the $RGB$ color space, and $I_x$, $I_y$, $I_{xx}$, $I_{yy}$ respectively corresponding to the first and second-order derivatives of the image intensities, estimated via the filters $[-1\ 0\ 1]$ and $[-1\ 2\ -1]$ in horizontal and vertical directions.

Next we group the image pixels into perceptually meaningful atomic regions using the SLIC superpixel algorithm [1]. The motivation behind this step is two folds. First, we use mean color of each foreground or background superpixel to reduce the sample space over which the representative samples are determined. Second, extracting these

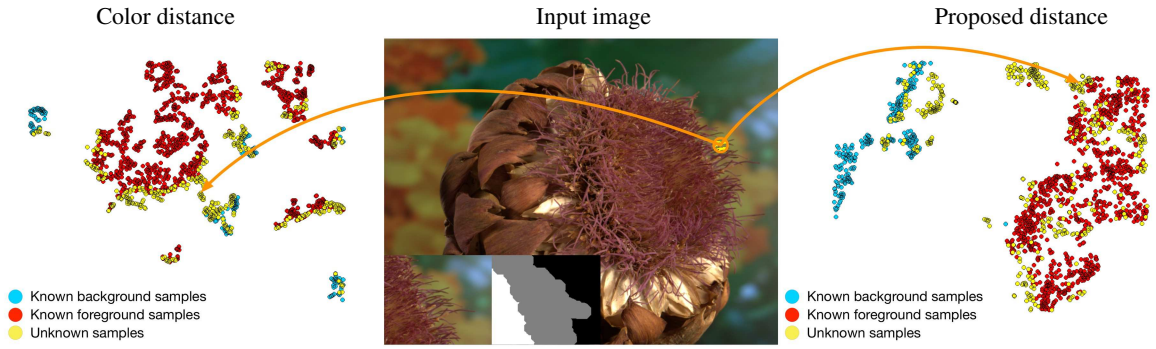



Figure 2. Distance embedding visualizations using t-SNE method [24] clearly demonstrate that the proposed KL-divergence based distance provides a better discrimination between known foreground and background pixels than using the standard color distance.

superpixels helps us to describe a pixel by means of the characteristics of its neighboring pixels, which provides a source of contextual information.

Let $s_p$ and $s_q$ respectively denote two superpixels. Then, one can use the KL-divergence to measure the distance between $s_p$ and $s_q$ by considering the corresponding feature distributions $P$ and $Q$ as

$$D_{KL}\left(P\|Q\right) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} \mathrm{d}x \tag{4}$$

In our formulation, we assume that each feature distribution can be modelled through a multivariate normal distribution such that $P \sim \mathcal{N}_{s_p} = \mathcal{N}\left(\mu_p, \Sigma_p\right)$. Here, $\mu_p$ denotes the 9-dimensional mean feature vector, and $\Sigma_p$ represents the $9 \times 9$ dimensional covariance matrix of features $\phi$ of the superpixel $s_p$. Then, the KL-Divergence between two superpixels $s_p$ and $s_q$ is described as follows:

$$\begin{aligned} D_{KL}(\mathcal{N}_{s_p}\|\mathcal{N}_{s_q}) = & \frac{1}{2}\Bigg(\mathrm{tr}\left(\Sigma_q^{-1}\Sigma_p\right) + \ln\left(\frac{\det\Sigma_q}{\det\Sigma_p}\right) \\ & + (\mu_q-\mu_p)^\top \Sigma_q^{-1}(\mu_q-\mu_p) - k\Bigg) \end{aligned} \tag{5}$$

with $k = 9$ denoting our feature dimension.

Note that the KL-divergence is not symmetric, hence we symmetrize it as follows to obtain a distance metric:

$$dist(s_p, s_q) = D_{KL}(\mathcal{N}_{s_p}\|\mathcal{N}_{s_q}) + D_{KL}(\mathcal{N}_{s_q}\|\mathcal{N}_{s_p}) \tag{6}$$

In measuring the distance between two superpixels $s_p$ and $s_q$, we found that, instead of using the metric in Eq. (6), the similarity and distance measures derived below lead to better discrimination:

$$S(s_p, s_q) = \frac{1}{dist(s_p, s_q) + \epsilon} \tag{7}$$

$$d(s_p, s_q) = 1 - S(s_p, s_q) \tag{8}$$

where we take $\epsilon = 0.5$ in the experiments.

In Figure 2, we qualitatively verify the effectiveness of our statistical distance measure over using only the mean color values of the superpixels. For a given input image, we compute the pairwise distances between the superpixels extracted from the known foreground and background, and unknown regions and then these distances are projected to a 2-dimensional space using t-SNE [24]. As can be seen, the proposed KL-divergence based distance measure provides better discrimination than simply using color distance.

### 2.2. Sampling via Sparse Subset Selection

Our strategy to obtain representative samples of known foreground and background regions to encode unknown region is inspired by the recently proposed Dissimilarity-based Sparse Subset Selection (DS3) algorithm [7], which formulate subset selection as a row-sparsity regularized trace minimization problem and presents a convex optimization framework to solve it. Suppose we use $\mathbf{K}$ and $\mathbf{U}$ to represent the set of superpixels extracted from the known foreground ($f$) and background ($b$), and unknown ($u$) regions, with $N = N_f + N_b$ and $M$ elements, respectively:

$$\begin{aligned} \mathbf{K} &= \{s_1^f, ..., s_{N_f}^f, s_1^b, ..., s_{N_b}^b\} \\ \mathbf{U} &= \{s_1^u, ..., s_M^u\} \end{aligned} \tag{9}$$

Assume that the pairwise dissimilarities $\{d_{ij}\}_{i=1,...,N}^{j=1,...,M}$ between superpixels of known region $\mathbf{K}$ and unknown region $\mathbf{U}$ are computed using the dissimilarity measure defined in Eq. (8)[1], and arranged into a matrix form as

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}_1^\top \\ \vdots \\ \mathbf{d}_N^\top \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1M} \\ \vdots & \vdots & & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{NM} \end{bmatrix} \in \mathbb{R}^{N\times M} \tag{10}$$

[1]We note that the approach is quite general in that it could work with dissimilarities which are asymmetric or violate the triangle inequality.

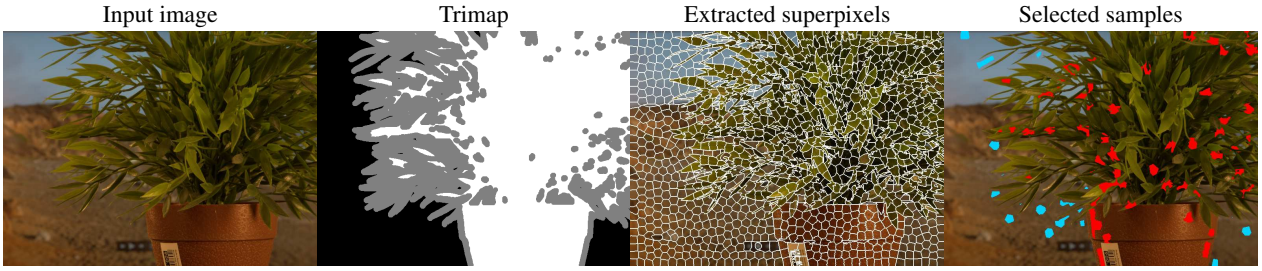


Figure 3. Sampling via sparse subset selection. Candidate foreground and background samples are shown in red and blue, respectively.

where the entries $d_{ij}$ signifies how well the superpixel $i$ represents the superpixel $j$, the smaller the value, the higher the degree of representativeness.

According to the method described in [7], in order to find a sparse set of samples of $\mathbf{K}$ that well represents $\mathbf{U}$, one can introduce a matrix of variables $\mathbf{P} \in \mathbb{R}^{N \times M}$ as

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_1^\top \\ \vdots \\ \mathbf{p}_N^\top \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ \vdots & \vdots & & \vdots \\ p_{N1} & d_{N2} & \cdots & p_{NM} \end{bmatrix} \tag{11}$$

whose each entry $p_{ij} \in [0, 1]$ is associated to $d_{ij}$ and denote the probability of superpixel $i$ being a representative for superpixel $j$. Then, the problem can be formulated as the following trace minimization problem regularized by a row-sparsity term:

$$\begin{aligned} \min_{\mathbf{P}} \quad & \gamma \, \|\mathbf{P}\|_{1,\infty} + \mathrm{tr}(\mathbf{D}^\top \mathbf{P}) \\ \text{s.t.} \quad & \mathbf{1}^\top \mathbf{P} = \mathbf{1}^\top, \mathbf{P} \geq 0 \end{aligned} \tag{12}$$

where the first term $\|\mathbf{P}\|_{1,\infty} \triangleq \sum_i \|\mathbf{p}_i\|_\infty$ penalizes the size of the representative set, the second term $\mathrm{tr}(\mathbf{D}^\top \mathbf{P}) = \sum_{ij} d_{ij} p_{ij}$ simply measures the total encoding cost, and the parameter $\gamma$ provides a trade-off between number of samples and encoding quality where smaller values of $\gamma$ will lead to less number of representative samples. An optimal solution $\mathbf{P}^*$ can be found very efficiently using an Alternating Direction Method of Multipliers (ADMM) approach [7], in which the indices from the nonzero rows of the solution $\mathbf{P}^*$ give us the selected samples of foreground and background superpixels, where we use the mean colors of these superpixels as the candidate set of foreground $F$ and background $B$ colors.

Figure 3 shows the samples obtained with our sparse sampling strategy on an illustrative image. As it can be seen, the proposed approach allows robust selection of a small set samples from the known regions where the selected samples are the samples amongst the ones that best represent the unknown regions. Hence, as compared to the existing sampling based models, we employ less number of samples to determine the alpha matte values of the unknown pixels.

### 2.3. Selecting The Best $(F, B)$ Pair

As compared to local sampling methods for image matting, which only collect samples near a given unknown pixel, employing a global scheme, such as ours, has the advantage of not missing any true samples if they are not located in the vicinity of the unknown pixel. In some cases, however, there is also a possibility that a local analysis may work better, especially when local samples are more strongly correlated with the unknown pixel. Hence, to get the best of both worlds, we decide to combine our global sparse sampling strategy with a local sampling scheme. Specifically, for a given unknown pixel, we enlarge the global candidate set to include 10 additional foreground and background samples which are selected from the spatially nearest boundary superpixels.

Once candidate foreground and background colors are sampled for an unknown pixel, we select the best foreground and background pair $(F, B)$ and accordingly determine its alpha matte value. In order to identify the best pair, we define a goodness function that depends on four different measures, which are described in detail below. In particular, in our formulation, we adopt the previously suggested chromatic distortion $C_u$ and spatial distance $S_u$ measures [12, 20, 21, 13] and additionally propose two new contextual similarity measures $T_u$ and $R_u$ to better deal with color ambiguity.

For an unknown pixel $u$ and a foreground-background pair $(F_i, B_i)$, the chromatic distortion $C_u$ measures how well the alpha matte $\hat{\alpha}$ estimated via Eq. (2) from $(F_i, B_i)$ fit to the linear composite equation given by Eq. (1), and is formulated as

$$C_u(F_i, B_i) = \exp(-\|I_u - (\hat{\alpha} F_i + (1 - \hat{\alpha}) B_i)\|) \tag{13}$$

where $I_u$ denote the observed color of the unknown pixel $u$.

The spatial distance measure $S_u$ quantifies the spatial closeness of the unknown pixel $u$ to the sample pair $(F_i, B_i)$ according to the distance between the coordinates of these pixels. Therefore, it favors selecting samples that are spatially close to the unknown pixel. It is simply defined as

$$S_u(F_i, B_i) = \exp\left(-\frac{\|u - f_i\|}{Z_F}\right) \cdot \exp\left(-\frac{\|u - b_i\|}{Z_B}\right) \tag{14}$$

where $f_i$ and $b_i$ respectively denote the spatial coordinates of the centers of the superpixels that are associated with the foreground and the background samples $F_i$ and $B_i$. The scalars $Z_F = (1/n_F)\sum_{k=1}^{n_F}\|u - f_k\|$ and $Z_B = (1/n_B)\sum_{k=1}^{n_B}\|u - b_k\|$ are used as scaling factors, which correspond to the mean spatial distance from the unknown pixel $u$ to all foreground samples $F$ with $n_F$ elements and all background samples $B$ with $n_B$ elements, respectively.

One of the great challenges in image matting is the color ambiguity problem which arises when the foreground and background have similar colors. As most of the matting studies consider pixel based similarities in comparing samples, they generally fail to resolve this ambiguity and incorrectly recognize an unknown foreground pixel as background or vice versa. To account for this, we introduce the following two additional local contextual similarity measures $T_u$ and $R_u$, which both exploit the similarity function defined in Eq. (7).

The first measure $T_u$ specifies the compatibility of the unknown pixel with the selected foreground and background samples, computed by means of their statistical feature similarities, and it provides a bias towards those pairs $(F_i, B_i)$ that have local contexts similar to that of the unknown pixel, and is formulated as

$$T_u(F_i, B_i) = S(s_{F_i}, s_u) + S(s_{B_i}, s_u) \tag{15}$$

where $s_{F_i}$, $s_{B_i}$, and $s_u$ respectively denote the superpixels associated with the corresponding foreground and background samples and the unknown pixel.

The second measure $R_u$ corresponds to a variant of the robustness term in [28], which builds upon the assumption that for any mixed pixel, the true background and foreground colors have similar feature statistics, calculated over the corresponding superpixels. Thus, it favors the selection of the foreground and the background samples that have similar contexts, and is defined as

$$R_u(F_i, B_i) = S(s_{F_i}, s_{B_i}). \tag{16}$$

Putting these four measures together, we arrive at the following objective function to determine the best $(F, B)$ pair:

$$\begin{aligned} O_u(F_i, B_i) = C_u(F_i, B_i)^c \cdot S_u(F_i, B_i)^s \cdot \\ T_u(F_i, B_i)^t \cdot R_u(F_i, B_i)^r, \end{aligned} \tag{17}$$

where $c, s, t, r$ are weighting coefficients, representing the contribution of the corresponding terms to the objective function. Empirically, we observed that that the color distortion $C_u$ and the contextual similarity measure $T_u$ are more distinguishing than others, and thus we set the coefficients as $c = 2, s = 0.5, t = 1, r = 0.5$.

### 2.4. Pre- and Post-Processing

Motivated by recent sampling based matting studies [21, 13], we apply some pre- and post-processing steps. First, before selecting the best $(F, B)$ sample pairs, we expand known regions to unknown regions by adopting the preprocessing step used in [21, 13]. Specifically, we consider an unknown pixel $u$ as a foreground pixel if the following condition is satisfied for a foreground pixel $f \in F$:

$$(D(I_u, I_f) < E_{thr}) \wedge (\|I_u - I_f\| \leq (C_{thr} - D(I_u, I_f))), \tag{18}$$

where $D(I_u, I_f)$ and $\|I_u - I_f\|$ are the spatial and the chromatic distances between the pixels $u$ and $f$, respectively, and $f$, and $E_{thr}$ and $C_{thr}$ are the corresponding thresholds which are all empirically set to 9. Similarly, an unknown pixel $u$ is taken as a background pixel if a similar condition is met for a background pixel $b \in B$.

Second, as a post-processing, we perform smoothing on the estimated alpha matte by adopting a modified version of the Laplacian matting model [16] as suggested in [9]. That is, we determine the final alpha values $\alpha^*$ by solving the following global minimization problem:

$$\begin{aligned} \alpha^* = \arg\min_{\alpha} \alpha^\top L\alpha + \lambda(\alpha - \hat{\alpha})^\top \Lambda(\alpha - \hat{\alpha}) \\ + \delta(\alpha - \hat{\alpha})^\top \Delta(\alpha - \hat{\alpha}) \end{aligned} \tag{19}$$

where the data term imposes the final alpha matte to be close to the estimated alpha matte $\hat{\alpha}$ from Eq. (2), and the matting Laplacian $L$ enforces local smoothing. The diagonal matrix $\Lambda$ in the first data term is defined using the provided trimap such that it has values 1 for the known pixels and 0 for the unknown ones. The scalar $\lambda$ is set to 100 so that it ensures no smoothing is applied to the alpha values of the known pixels. The second diagonal matrix $\Delta$, on the other hand, is defined by further considering the estimated confidence scores in a way that it has values 0 for the known pixels and the corresponding confidence values $O_u(F, B)$ from Eq. (17) for the unknown pixels. The scalar $\delta$ here is set to 0.1 and determines the relative importance of the smoothness term which considers the correlation between neighboring pixels.

## 3. Experimental Results

We evaluate the proposed approach on a well-established benchmark dataset [19], which contains 35 natural images, each having a foreground object with different degrees of translucency or transparency. Among those images, 27 of them constitute the training set where the groundtruth alpha mattes are available. On the otherhand, the remaining 8 images are used for the actual evaluation, whose groundtruth alpha mattes are hidden from the public to prevent parameter tuning. In addition, for each test image, there are three matting difficulty levels that respectively correspond to small, large and user trimaps. To quantitatively evaluate our approach, in the experiments, we consider three different metrics, namely, the mean square error (MSE), the sum of absolute differences (SAD) and the gradient error.

Table 1. Evaluation of matting methods the benchmark dataset [19] with three trimaps according to SAD, MSE and Gradient error metrics.

| Sum of Absolute Differences | | | | | Mean Square Error | | | | | Gradient Error | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | overall rank | avg. small rank | avg. large rank | avg. user rank | Method | overall rank | avg. small rank | avg. large rank | avg. user rank | Method | overall rank | avg. small rank | avg. large rank | avg. user rank |
| 1. LNSP Matting | **7.0** | 4.5 | 6.1 | 10.4 | 1.LNSP Matting | **6.5** | 4.6 | 5.4 | 9.5 | 1. Proposed Method | **7.5** | 6.6 | 5.9 | 10.0 |
| 2. Proposed Method | **7.6** | 7.5 | 6.1 | 9.3 | 2. Proposed Method | **7.9** | 7.8 | 6.3 | 9.6 | 2. LNSP Matting | **7.9** | 6.3 | 6.9 | 10.6 |
| 3. Iterative Transductive | **8.7** | 10.1 | 7.9 | 8.1 | 3. CCM | **8.0** | 10.3 | 7.8 | 6.1 | 3. Comprehensive sampling | **8.3** | 8.5 | 7.6 | 8.8 |
| 4. Comprehensive sampling | **8.8** | 7.4 | 8.6 | 10.5 | 4. Comprehensive sampling | **8.7** | 7.9 | 8.6 | 9.5 | 4. CCM | **9.8** | 12.0 | 8.9 | 8.5 |
| 5. CWCT sampling | **9.4** | 9.9 | 9.6 | 8.8 | 5. SVR Matting | **9.5** | 12.1 | 8.0 | 8.4 | 5. SVR Matting | **9.9** | 11.8 | 10.5 | 7.4 |
| 6. SVR Matting | **9.7** | 11.8 | 9.0 | 8.4 | 6. CWCT sampling | **9.8** | 9.9 | 10.4 | 9.3 | 6. Sparse coded matting | **10.3** | 11.6 | 9.0 | 10.3 |
| 7. Sparse coded matting | **9.9** | 12.0 | 10.3 | 7.5 | 7. Sparse coded matting | **11.8** | 13.4 | 12.4 | 9.6 | 7. Segmentation-based | **10.6** | 13.5 | 8.8 | 9.5 |
| 8. WCT sampling | **10.8** | 9.0 | 12.3 | 11.0 | 8. WCT sampling | **11.9** | 10.6 | 12.9 | 12.1 | 8. Global Sampling | **10.7** | 10.9 | 11.1 | 10.1 |
| 9. CCM | **11.0** | 13.0 | 10.4 | 9.5 | 9. Global Sampling | **12.4** | 8.8 | 15.1 | 13.3 | 9. Shared Matting | **10.9** | 11.0 | 11.4 | 10.4 |
| 10. Shared Matting | **12.1** | 11.4 | 14.4 | 10.5 | 10. Iterative Transductive | **12.9** | 14.0 | 11.5 | 13.3 | 10. Improved color matting | **11.2** | 12.6 | 11.5 | 9.4 |

**Effect of $\gamma$ parameter.** Fig. 4 shows that the average MSE values over all the training images and all trimaps do not vary much for different values of $\gamma$. These results seem to be consistent with the theoretical analysis in [7] that for a proper range of values, the DS3 algorithm that we utilize in sampling is guaranteed to find representative samples from all groups when there is a mutual relationship between known and unknown sets. In the remaining experiments, $\gamma$ is set to 0.025 as it provides the minimum MSE value for the training set.

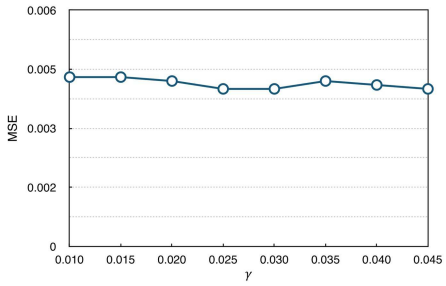



Figure 4. Effect of $\gamma$ parameter on the performance. Plot shows average MSE values over all training images and all trimaps.

**Effect of local samples.** In Fig. 5, we show the effect of including local samples from boundary to the candidate set found by the proposed sparse sampling scheme. The numbers in the plot refer to the MSE errors averaged over all test images. For each trimap type, adding some closest boundary pixels further improves the performance. The smallest gain is in the large trimaps since having more number of known pixels helps our sparse sampling method to better exploit the associations between the known and unknown regions, eliminating the need for local samples.

**Comparison with the state-of-the-art.** Table 1 presents the quantitative comparison of our approach and nine best performing matting algorithms on the benchmark hosted at *www.alphamatting.com* [19] where we report the average rankings over the test images according to SAD, MSE and gradient metrics for all three different types of trimap, and the overall ranks, computed as the average over all the images and for all the trimaps. Overall, our approach provides

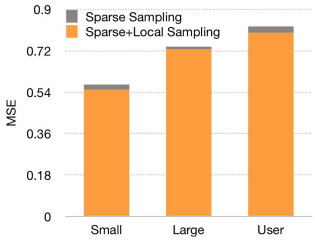
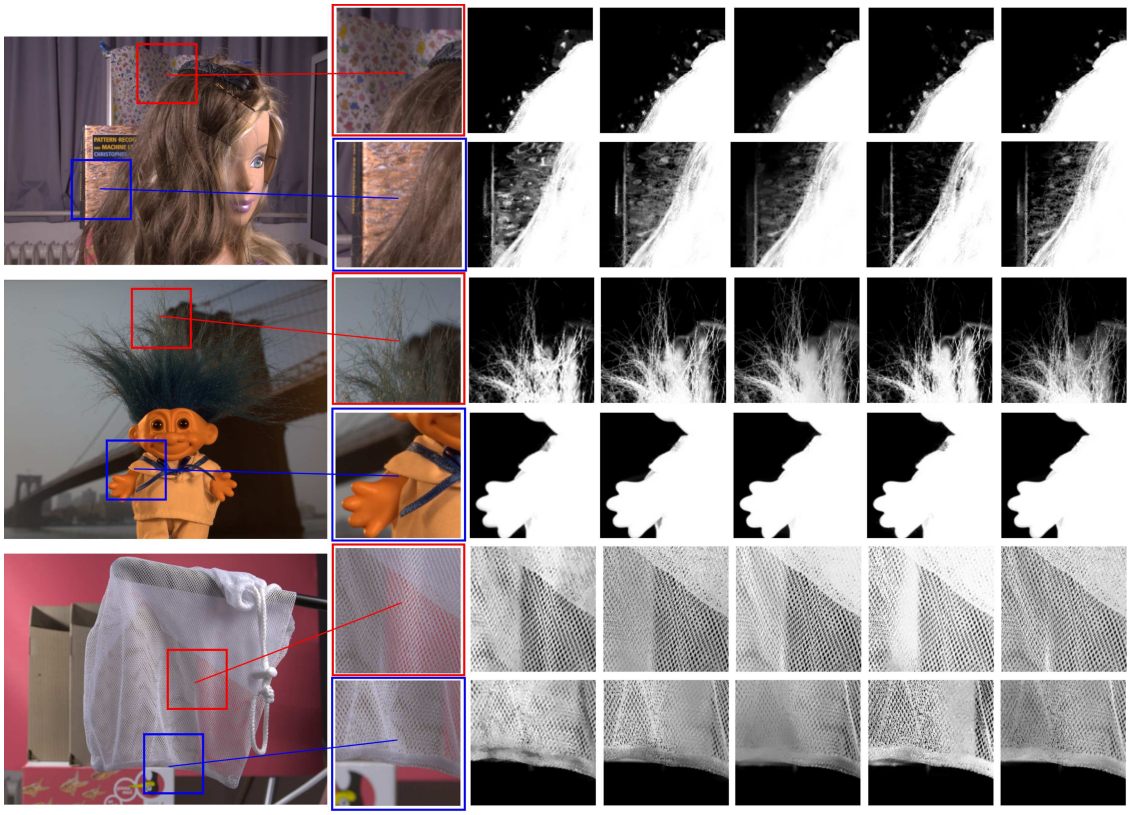



Figure 5. Effect of including local samples to the representative set obtained with the proposed sparse sampling scheme. Plot shows average MSE values over all test images for three types of trimaps.

highly competitive results against the state-of-the-art methods. It ranked the best with respect to the gradient error and the second best for the other two metrics. Especially, it outperforms all the existing sampling-based matting methods. Fig. 3 provides qualitative comparisons of our approach and the recent matting studies [4, 25, 21, 13] on the *doll*, *troll* and *net* images from the benchmark dataset. Additional results are provided as supplementary material.

*Textured background.* In the first row of Fig. 3, we show the ability of our approach to naturally handle textured backgrounds via the proposed KL-divergence based contextual measure. For the *doll* placed in front of a highly textured background, while other matting methods, including CWCT sampling [25] which employs an additional texture compatibility measure, tend to interpret some of the colored blobs in the background as foreground, our model produces a much more accurate alpha map.

*Color ambiguity.* When the foreground object and the background have similar color distributions, most of matting studies suffer from the so-called color ambiguity and fail to provide reliable alpha values for the unknown pixels. Both second and third rows of Fig. 3 illustrate this issue where the colors of the book and the bridge in the background is very similar to those of the hairs of the *doll* and *troll*, respectively. For these examples, CWCT [25] and Comprehensive sampling [21] give inaccurate estimations whereas LNSP matting [4] oversmooths the foreground matte. Sparse coded matting [13] provides better results but misses some of the foreground details in the hairs. On the other hand, our method is able to achieve significantly bet-



(a) (b) (c) (d) (e) (f)

Figure 6. Visual comparison of our approach with other sampling-based image matting methods. (a) Input image, (b) CWCT sampling [25], (c) Comprehensive sampling [21], (d) LNSP matting [4], (e) Sparse coded matting [13] and (f) Proposed approach.

ter results, providing a more robust discrimination between the background and the foreground.

*Missing samples.* Previously proposed sampling-based matting methods typically employ certain assumptions while collecting samples from known regions but these assumptions might sometimes lead to missing true foreground and background colors for some unknown pixels. In the fourth row of Fig. 3, we demonstrate the effectiveness of our sparse sampling strategy on the *troll* image. While the other sampling based methods [25, 21, 13] incorrectly recognize the blue ribbon as mixed pixels, our algorithm successfully interprets it as a part of the foreground object. Likewise, the LNSP matting [4] produces an alpha map similar to ours as it uses a non-local smoothness prior in their formulation.

*Translucent foreground.* Transparent or translucent objects pose another great challenge for matting as they make collecting true foreground color samples difficult. The last two rows of Fig. 3 shows the matting results of two different regions from the *net* image in detail where such a foreground object exists. Due to the characteristics of the test image, all of the competing matting methods fail to differentiate background pixels from the foreground although the distributions of the background and foreground colors are well separated. In contrast, our approach produces a remarkably superior alpha matte.

We note that our approach also works with sparse user inputs. Some sample results with sparse scribbles and our runtime performance are presented in our supp. material.

## 4. Conclusion

We developed a new and theoretically well-grounded sampling strategy for image matting. Rather than making assumptions about the possible locations of true color samples, or performing a direct clustering of all known pixels, our sampling scheme solves a sparse subset selection problem over known pixels to obtain a small set of representative samples that best explain the unknown pixels. Moreover, it employs a novel KL-divergence based contextual measure in both collecting the candidate sample set and finding the best $(F, B)$ pair for an unknown pixel. Our experiments clearly demonstrate that our approach is superior to existing sampling-based image matting methods and achieves state-of-the-art results.

## Acknowledgments

This work was supported by the Scientific and Technological Research Council of Turkey – Award 112E146.

## References


[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11):2274–2282, 2012. 3

[2] A. Berman, A. Dadourian, and P. Vlahos. Method for removing from an image the background surrounding a selected object, Oct. 17 2000. US Patent 6,134,346. 2

[3] Q. Chen, D. Li, and C.-K. Tang. Knn matting. In *CVPR*, pages 869–876, 2012. 1

[4] X. Chen, D. Zou, S. Z. Zhou, Q. Zhao, and P. Tan. Image matting with local and nonlocal smooth priors. In *CVPR*, pages 1902–1907, 2013. 1, 7, 8

[5] I. Choi, M. Lee, and Y.-W. Tai. Video matting using multi-frame nonlocal matting laplacian. In *Computer Vision–ECCV 2012*, pages 540–553. Springer, 2012. 3

[6] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *CVPR*, volume 2, pages II–264, 2001. 1

[7] E. Elhamifar, G. Sapiro, and S. S. Sastry. Dissimilarity-based sparse subset selection. *arXiv preprint arXiv:1407.6810*, 2014. 2, 3, 4, 5, 7

[8] E. Elhamifar, G. Sapiro, and R. Vidal. Finding exemplars from pairwise dissimilarities via simultaneous sparse recovery. In *Advances in Neural Information Processing Systems*, pages 19–27, 2012. 2

[9] E. S. Gastal and M. M. Oliveira. Shared sampling for real-time alpha matting. *Computer Graphics Forum*, 29(2):575–584, 2010. 1, 2, 3, 6

[10] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann. Random walks for interactive alpha-matting. In *Int. Conf. Vis., Imag., Image Process.*, pages 423–429, 2005. 1

[11] B. He, G. Wang, C. Shi, X. Yin, B. Liu, and X. Lin. Iterative transductive learning for alpha matting. In *ICIP*, pages 4282–4286, 2013. 1

[12] K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun. A global sampling method for alpha matting. In *CVPR*, pages 2049–2056, 2011. 1, 2, 3, 5

[13] J. Johnson, D. Rajan, and H. Cholakkal. Sparse codes as alpha matte. In *BMVC*, 2014. 1, 3, 5, 6, 7, 8

[14] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, pages 79–86, 1951. 3

[15] P. Lee and Y. Wu. Nonlocal matting. In *CVPR*, pages 2193–2200, 2011. 1

[16] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):228–242, 2008. 1, 6

[17] Y. Mishima. Soft edge chroma-key generation based upon hexoctahedral color space, Oct. 11 1994. US Patent 5,355,174. 2

[18] C. Rhemann, C. Rother, and M. Gelautz. Improving color modeling for alpha matting. In *BMVC*, 2008. 2

[19] C. Rhemann, C. Rother, J. Wang, M. Gelautz, P. Kohli, and P. Rott. A perceptually motivated online benchmark for image matting. In *CVPR*, pages 1826–1833, 2009. 3, 6, 7

[20] E. Shahrian and D. Rajan. Weighted color and texture sample selection for image matting. In *CVPR*, pages 718–725, 2012. 1, 2, 3, 5

[21] E. Shahrian, D. Rajan, B. Price, and S. Cohen. Improving image matting using comprehensive sampling sets. In *CVPR*, pages 636–643, 2013. 1, 2, 3, 5, 6, 7, 8

[22] Y. Shi, O. C. Au, J. Pang, K. Tang, W. Sun, H. Zhang, W. Zhu, and L. Jia. Color clustering matting. In *ICME*, pages 1–6, 2013. 1

[23] J. Sun, J. Jia, C.-K. Tang, and H.-Y. Shum. Poisson matting. *ACM Trans. Graph.*, 23(3):315–321, 2004. 1

[24] L. van der Maaten and G. Hinton. Visualizing high-dimensional data using t-SNE. *J. Machine Learning Research*, 9:2579–2605, Nov 2008. 4

[25] E. Varnousfaderani and D. Rajan. Weighted color and texture sample selection for image matting. *IEEE Trans. Image Processing*, 22(11):4260–4270, Nov 2013. 1, 3, 7, 8

[26] J. Wang and M. Cohen. Image and video matting: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3(2):97–175, 2007. 1

[27] J. Wang and M. F. Cohen. An iterative optimization approach for unified image segmentation and matting. In *ICCV*, volume 2, pages 936–943, 2005. 1

[28] J. Wang and M. F. Cohen. Optimized color sampling for robust matting. In *CVPR*, pages 1–8. IEEE, 2007. 2, 3, 6

[29] Z. Zhang, Q. Zhu, and Y. Xie. Learning based alpha matting using support vector regression. In *ICIP*, pages 2109–2112, 2012. 1

[30] Q. Zhu, L. Shao, X. Li, and L. Wang. Targeting accurate object extraction from an image: A comprehensive study of natural image matting. *IEEE Trans. Neural Networks and Learning Systems*, 26(2):185–207, 2015. 1