SSURJET

Sir Syed University Research Journal of Engineering & Technology
2023, Vol. 13, No. 2,
https://doi.org/10.33317/ssurj.564

Image Matting using Superpixels Centroid

Anam Akbar[1], Aniqa Shirazi[1], and Muhammad Sarim Farooqui[1]

[1]*Department of Computer Science, Federal Urdu University for Arts Science and Technology, Karachi, Pakistan*

Correspondence Author: Anam Akbar (a.akbar@fuuast.edu.pk)

Received February 11, 2023; Revised June 06, 2023; Accepted June 21, 2023

Abstract

The orientation and focus of this research piece are the extraction of the foreground and the compositing of this extracted region into another background. This phenomenon is termed as, 'Image Matting', which is frequently employed in film production or the digital media world. The proposed method approaches the ill-posed nature of image matting via a non-parametric sampling-based method along with the clustering technique known as 'Superpixel'. In the proposed method, pixels of the entire image(s) tend to gather in close proximity under one unit (Superpixel) with respect to color, intensity, and texture. This gathering in close proximity reduces the search space more than 20 times and helps in efficiently finding the association of unknown regions with the samples from the background and foreground. The use of samples facilitates the pixel color assimilating with local image structure, which is significant to calculate a good resultant alpha matte particularly in the image having complex texture, and in natural images. As per my knowledge and study, the matting problem using centroids of Superpixels has not previously been explored. Results are evaluated on different images on an online standard open-source dataset, available for image matting. Results are comparable to the different matting algorithms applied independently on images of the dataset. Confidence and ameliorated refinements of proposed methods are evident in the obtained results compared to other image matting methods.

Index Terms: *Alpha Matte, Global/Local Samples, Image Matting, Non-Parametric, Super-Pixel.*

I. INTRODUCTION

Image matting is best described as the accurate and exact extraction of the foreground region from a composite image. Composite image 'I_x' is the combination of foreground pixels 'F_x', background pixels 'B_x', and alpha matte. In computer vision; image matting plays an imperative role in media production and video editing especially when there is a need to combine a foreground object into another different desired background scene. Due to the Complex nature of image and video matting various researchers have marked their contribution in matting literature and several are in progress. The idea of channel separation along with all color dimensions, to preserve the opaqueness details of foreground objects in a site was brought forward by Alvy Ray Smith [1] in 1970. Later Porter and Duff [2] developed the idea in their significant research paper for compositing digital image(s). They proposed an important equation, i.e., eq. (1) known as the compositing equation and formulated that governs digital image matting.

$$I_x = \alpha_x F_x + (1 - \alpha_x) B_x \quad (1)$$

The concern about matting is that, it is an ill-posed method, as there are three numbers of unknown elements; the first is eq. (1). While $\boldsymbol{\alpha_x}$ is between 0 and 1; where 1 means the imperfect-transparent foreground pixels. If pixel x is said to be a distinct foreground if alpha is 1 and a distinct background if alpha is 0; if it is not then it is defined as mixed.

Various research methods that approach the image matting problem, segregate whole image pixels as exact foreground, mixed, and exact background with the help of trimap [3]. To decide mixed pixels as foreground and background alpha values each pixel is estimated. The under-constrained matting problems are generally approached in two interactive ways, Scribble-based [4], and [5], and Trimap-based image matting. The defined matting problem can be minimized with the help of a trimap. Figure I trimap is a type of input that corresponds to its original image taken from our reference i.e., [6].

Trimap can be generated by different segmentation algorithms or with the help of user manual interaction [7], and [8] using trimap as an input can produce desirable and optimal results. Hard segmentation and soft segmentation are correlated to each other in this process, to correctly take out semantically significant foreground regions. Approximately, all image matting techniques begin with an input image based on three important regions: foreground background, and unknown. This triad input pixel map is termed as trimap as shown in Figure 1. Thus the matting problem is minimized to reckon and wring foreground pixels ($\boldsymbol{F_x}$) background pixels ($\boldsymbol{B_x}$) and ($\boldsymbol{\alpha_x}$) for pixels in the grey region. Existing matting techniques can be segregated into three general categories affinity-based methods and sampling-based methods.

Affinity-based approaches assume region-based image statistics by considering different affinities between neighbor region pixels. As a replacement for straight estimating the alpha value at each and every pixel, it basically models the matte gradient across the image lattice. The pixel correlations are typically well-built so the local smoothness supposition normally holds, still for moderately multifarious images. In Poisson matting [9] the assumption is over locally smooth of an intensity change.

Random Walk matting [10] proposed a similar method to Poisson matting that is based on the affinity matrix in the definition of the Gaussian function. In Closed Form matting [11] a newly defined quadratic cost function is formulated in 'α_x' to estimate the optimum alpha matte. KNN matting [12] assumes non-local principles with K nearest neighbors. Since pixel matting statistics in these approaches are transmitted from the identified known regions to the unidentified unknown regions therefore leads to the error across the alpha matte.

Figure I: One of Dataset Representation Input Image with its Trimap Provided by Alpha Matting Evaluation Website [6]

Generally, sampling-based methods find the relation of alpha parameters with the local samples [7], and [8]. Such methods gather a set of samples from known regions (foreground/ background) to calculate the alpha values of unknown pixels. Different techniques are practiced to identify the samples from known regions that optimally represent the correct foreground and background colors of every unknown pixel in the Trimap. When the optimum known foreground and background samples are identified for every pixel x, its alpha value is estimated as eq. (1). This approach can further be divided into two parts parametric and nonparametric. Parametric methods [13-15] sample of foreground and background regions are selected, this method generally fits parametric statistical models to the background and foreground samples, normally Gaussians distribution. For estimating the alpha value of a selected pixel from an unknown region, find the distances of this pixel to the foreground and background distributions. Non-parametric methods [16-19] basically collect samples of known foreground (F) and known background (B) to estimate alpha values of pixels belonging to unknown regions, regardless of fitting parametric statistics. However, the estimated alpha matte is extremely relying on the chosen samples from the known region. It reduces the quality; if the desired color of the foreground and the desired color of the background in unknown pixels are not present in the sample sets that were selected. Mishima introduced blue screen matting [20], it also depends on foreground and background representative samples. The background has just one color bunch, each background pixel can be enclosed by a little globe approximated by a polyhedra (triangular mesh) in the color space. Each foreground pixel constructs other polyhedra outer the background. The alpha value of an unknown pixel is then estimated by compiling its relative position to the other polyhedra. The knockout technique estimates the unknown by taking the weighted sum of the close neighbor's pixels' color in the foreground region [21]. Weights are proportional to their local spatial distances to respective unknown pixels. The same method is followed to estimate background color. With respect to the relative position of these unknown pixels, their association with foreground and background is decided. Finally, in each separate color channel, all unknown mixed pixels are estimated three times w.r.t color channels. The Robust matting technique introduced the selection of confidence sample pairs [16]. This confidence is defined by color fitness criteria that take samples having a high rate of spatial resemblance in Euclidean space. This reduces the matting cost function. These high-order samples estimate both foreground and background color and later this information facilitates the computation of alpha matte. Improved color modeling method uses the same strategy but the selection of samples is based on geodesic distance instead of Euclidean distance [17]. A Shared matting algorithm gathers samples from the emanating border of mixed pixels' region., each unknown pixel collects very few samples, but the samples are further shared among neighboring pixels [18]. Selection of the sample is on the basis of probabilistic attributes, photometric attributes, and even spatial process attributes are considered. Approaches like [17], and [22-26] are worked on learning-based methods. The approaches combine the global and local results of KNN and Closed-form matting and regenerate alpha matte from the obtained results [8]. Further to gain more confidence in the alpha matte quality, the process proceeded with the Deep Convolutional Neural Network (DCNN). FBA-net is introduced by authors to approximate alpha with foreground and background [27]. Researchers have proposed a model that works on semantic and Textural Compensate Path (TCP) in parallel, A Feature Fusion Unit (FFU) fused multi-scale features from the semantic path and passed them to TCP [28].

In reference [29] non-negative matrix factorization is used to estimate alpha matte. It assumes two spatial patches. These approaches produce inconsistency in the resultant alpha matte due to local samples are not truly representative of unknown pixels. The method of Fast and Adaptive Trimaps (FATs) is introduced in [6]. Extraction of the foreground is obtained via superpixel and then the Grab cut method is applied to gain a raw mask. Existing matting algorithm are not consistent in terms of large sampling search but also lacks in smoothing and extracting desired alpha.

In this paper, a global sampling-based method is introduced to minimize the true samples problem in the no-parametric approach. A clustering-based technique called Superpixel is used to efficiently find the samples from known regions. The superpixel technique reduces the foreground and background known region more than 20 times so it can efficiently find the optimum samples in the background and foreground for every unknown pixel. A high-confidence foreground and background pair set of colors is selected using Euclidean distance analysis to estimate the alpha matte.

II. SUPERPIXEL

Superpixel is a method that perceptually sectionizes image pixels on the basis of color, intensity, and texture. Each atomic section facilitates delimiting of pixels' redundancy in the sample and provides primal to estimate image distinctive and resembled features. These methods have set their marks in many computer vision algorithms, such as object segmentation in images and videos.

Figure II demonstrates image segmentation into superpixels with distinct labels for each superpixel. The number of original unit samples is reduced by a considerable amount, from 76800 pixels to 2843 super-pixels. Every unit possesses similar properties in terms of texture, intensity, and color.

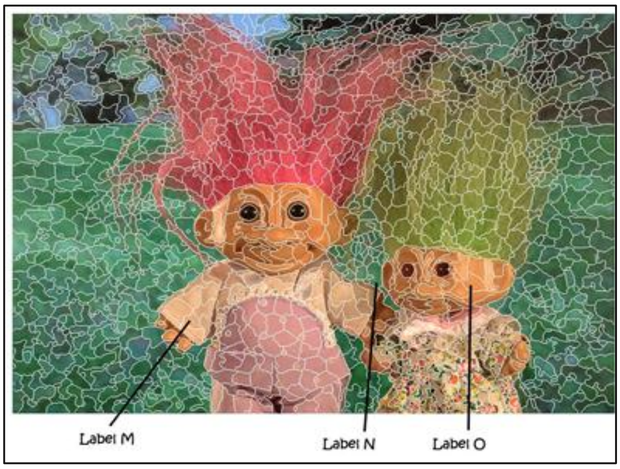

Figure II: Image Segmentation after Generating Superpixels using SLIC Algorithm with Distinct Labels

There are two categories of algorithms presented for generating superpixels; graph-based approach and gradient ascent approach. In Graph-based approaches [30-33] every pixel is assumed as a node like in a graph for generating superpixel. Similarities among neighboring pixels are calculated using edge weights between two nodes. Superpixels are generated by reducing a cost function described in the graph. In Gradient-ascent-based algorithms [26], [34], and [35], initialization starts from an uneven preliminary clustering of pixels, and gradient ascent techniques recursively filter the clusters until various convergence standard is achieved to generate superpixels. In the proposed research method a Simple Linear Iterative Clustering (SLIC) algorithm is used [36]. SLIC is a recent approach for producing superpixels that is more robust than existing superpixel algorithms, more memory efficient, exhibits traditional unit boundary attachment, and better optimizes the performance efficiency of segmentation algorithms. Simple Linear Iterative Clustering (SLIC) is an adaptation of k-means for generating superpixels with two significant distinctions. The first difference is the reduced distance calculation by delimiting the search space to a specific region which is proportional to the superpixel size. This minimizes the complexity in terms of pixels N and is free from the number of superpixels K. The Second difference is the combined measurement of superpixel closeness via weighted distance in both spatial and color spaces, while at the same time giving control over the superpixel's compactness and size.

In the proposed algorithm different algorithms [26], [30], and [34] for generating superpixels are tested and examined. In the proposed technique Simple Linear Iterative Clustering (SLIC) is used because it has found optimum results and it is easy to implement [37-39].

III. IMAGE MATTING USING SUPERPIXEL CENTROID

The proposed approach is initialized by a user-defined trimap. The proposed algorithm can be split into five main steps:

A. Applying Superpixels Technique on the Image along with Trimap

SLIC Superpixels technique is implemented in Matlab, shareware codes are available. Firstly, the SLIC technique is applied to the image along with Trimap. Firstly, applying SLIC on the natural image will return unique labels for each superpixel, here for generating the desired number of superpixels can be specified. Commission International de l'Eclairage (CIELAB) color model is used in this algorithm for color images, this procedure for clustering initiates with centers of k initial clusters are defined as $C_{m=[l_m\ a_m\ b_m\ \hat{x}_m\ \hat{y}_m\ ]^T}$ these are intended samples on a standard grid spaced Sm pixels separately. To generate superpixels approximately uniformly in size, the grid time period is Sm = $\sqrt{N/K}$. Then the centers of clusters are shifted to seed positions equivalent to the minimum gradient location in a size of 3 × 3 nearest neighborhood window. After performing the assignment, each pixel corresponds with the adjacent nearest center of the cluster whose search area overlies its place. Measurement of distance 'Dm' can be done as calculated in eq. (4). To combine the two distances first distance for CIELAB color space 'dcs' and the second distance 'dp' between positions of each pixel and cluster center Cm can be calculated under a single unit $\widehat{D}$ as eq. (2).

$$d_{cs} = \sqrt{(l_x - l_m)^2 + (a_x - a_m)^2 + (b_x - b_m)^2}$$

$$d_p = \sqrt{(\hat{x}_x - \hat{x}_m)^2 + (\hat{y}_x - \hat{y}_m)^2}$$

$$\widehat{D} = \sqrt{\left(\frac{d_{cs}}{N_{cs}}\right)^2 + \left(\frac{d_p}{N_p}\right)^2} \quad (2)$$

Where Ncs and Np are normalization to color and spatial proximity by their maximum distances within each cluster. The maximum spatial distance 'Np' estimated to the sampling period Sm =$\sqrt{N/K}$. The maximum color distance can vary from cluster to cluster so assumes its value assumes a constant m so eq. (2) can be written as eq. (3).

$$\widehat{D} = \sqrt{\left(\frac{d_{cs}}{m}\right)^2 + \left(\frac{d_p}{s}\right)^2} \quad (3)$$

It can simplify the actual distance measure which used in practice:

$$D_m = \sqrt{(d_{cs})^2 + \left(\frac{d_p}{S_m}\right)^2} \quad (4)$$

The measurement of a distance 'Dm' finds the nearest center of the cluster for every pixel. While the estimated spatial level of the calculated superpixel is an area of estimated size Sm × Sm, exploration for the same pixels is completed in an area of 2Sm ×2Sm around the center of each superpixel. When every pixel has been related to the nearest center of the cluster, a new step alters the centers of all clusters to the mean [l a b xˆyˆ]^T vector of entire pixels associating with the cluster. The L2 norm is applied to calculate a residual error E, this error is calculated between updated center positions of clusters and previous center positions of clusters. The allocation and update process can be iterated repeatedly until the error congregates. At last, a last step of post-processing is applied for connecting disjoint pixels by reassigning them to the nearest superpixels.

B. Collecting Superpixel as a Sample for Unknown

There are a number of superpixels, but in proposed algorithm neglected those superpixels that exist in both unknown and foreground regions and unknown and background regions. In trimap black portion indicates the background region (0 is for black), the White portion indicates the foreground region (255 is for Foreground) and the gray portion indicates the unknown region (128 is for gray) as shown in figure I. The distinction between pure foreground and background superpixels for samples as:

$L(i)$ == 128 ignore
$L(i)$ == 255 foreground samples
$L(i)$ == 0 background samples

Where '*L'* indicates the unique label of each superpixel '*I'* of every pixel. In figure III violet color highlighted area shows superpixels having unknown regions along with foreground and background. Superpixels having both background and unknown regions magnificently are shown in figure IV.

C. Selection of Best Superpixel using Centroid

In this step selection of the best sample pair of superpixel for each unknown using region property 'centroid' is performed. Superpixel has the same type of pixels in it so the centroid of a superpixel has approximately the same properties as other pixels in the same superpixel. Then find the centroid of all superpixels having only background pixels as shown in figure V red dots or pixels showing the centroid of background superpixels.

Figure III: Violet Color Showing Unknown Pixels, Neglecting Superpixels having Unknown

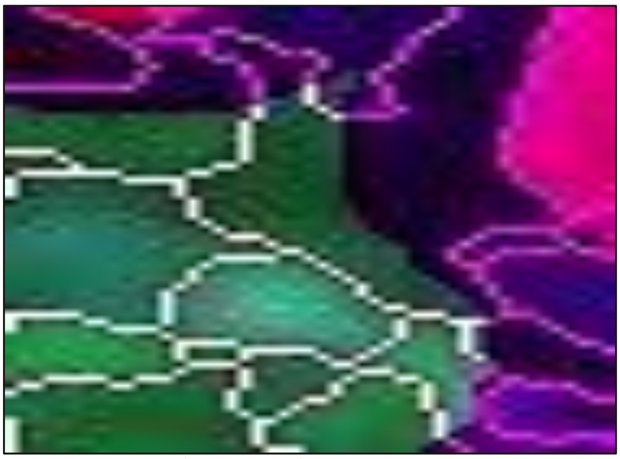

Figure IV: Magnificent View of Superpixels Having Two Regions

Figure V: Red Dots Showing Centroid of Pure Backgrounds

Figure VI, blue dots or pixels showing the centroid of foreground superpixels. After finding centroids the next step is to find the nearest superpixels in the background and foreground for every unknown pixel 'k' using equation of Euclidean distance formula i.e., eq. (5). Every superpixel centroid 'C' and unknown pixel 'U' have their Cartesian Coordinates, for finding the nearest superpixel for each unknown pixel the distance is calculated for each unknown pixel 'k' to each background centroid 'm' and each foreground centroid 'n' where found minimum distances for both foreground and background update the unique labels of background and foreground superpixels separately.

Figure VI: Blue Dots showing the Centroid of Pure Foreground's Superpixels

$$D = \sqrt{(C_{\acute{x}} - U_{\acute{x}})^2 + (C_{\acute{y}} - U_{\acute{y}})^2} \quad (5)$$

D. Optimizing of Alpha Matte

After selecting the best superpixels '*Lb*' of both background and foreground regions for each unknown pixel 'k', then select 'n' pixels from each selected superpixel as background color samples 'B^m' and foreground color samples 'F^n' for all unknown pixels' 'k' with corresponding to their background and foreground superpixels. After that leading objective is to calculate the most favorable foreground and background set from the samples set. The best foreground and background pair sets for each unknown pixel 'k' are calculated using eq. (6).

$$(\alpha\_k, F\_k, B\_k) = argmin\,\{C - [\alpha\hat{}u\, F\hat{}n + (1 - \alpha\hat{}u)\, B\hat{}m]\} \quad (6)$$

where 'αu' are different levels of alpha values starting from 0 to 1 with an increment of 0.1, and 'B^m' and 'F^n' are background and foreground color samples for each unknown pixel 'I' respectively. The above equation i.e., eq. (6) gives the optimum result of alpha, foreground, and background for every value of unknown pixel 'k'.

Algorithm 1: Selection of the Best Superpixel

```
/ * Initialization * /
Centroid values for all Background superpixels
    Cb(m) = [Cbx' Cby']
Centroid values for all Foreground superpixels
    Cf(n) = [Cfx' Cfy']
Pixel values for all unknown pixels
        U(k) = [Ux' Uy']
        where k = 1 x no. of unknown pixels
Labels of all Superpixels L(i)
for each unknown pixel k do
Set df(k) = ∞   distance for each unknown pixel k to foreground
Set db(k) = ∞   distance for each unknown pixel k to background
  for each center of foreground superpixels Cf(n) do
    Compute distance Db between Cb(m) and U(k)
    Db = sqrt((Cb_x̂(m) − U_x̂(k))^2 − (Cb_ŷ(m) − U_ŷ(k))^2)
    if Db < db then
        set db = Db
        Lb(k,1) = L(Cb)
    end if
  end for
  for each center of foreground superpixels Cf(n) do
    Compute distance Df between Cf(n) and U(k)
    Db = sqrt((Cf_x̂(n) − U_x̂(k))^2 − (Cf_ŷ(n) − U_ŷ(k))^2)
    if Df < df then
        set df = Df
        Lb(k,2) = L(Cf)
    end if
  end for
end for
```

E. Smooth the Alpha Matte

It is observed the quality of primarily estimated alpha matte needs additional improvement. This improvement can be done by a technique of matte optimization. Matting optimization considers correlation by calculating the relation among local pixels in neighboring regions. In this step matte optimization technique is applied as used in shared matting for further quality improvement of alpha matte as calculated using eq. (6). In shared matting [18] optimization data terms and smoothness terms determine the cost function. Like in Closed form matting [19], earlier estimated values of $'\alpha'$ along with the data term is used as smoothing term while matting Laplacian. The final optimum alpha is estimated as eq. (7).

$$\hat{\alpha} = argmin\ \hat{\alpha}^{\tau} L_c\, \hat{\alpha} + \lambda(\hat{\alpha} - \alpha)^{\tau} D_g (\hat{\alpha} - \alpha) + \gamma(\hat{\alpha} - \alpha)^{\tau} \Gamma\, (\hat{\alpha} - \alpha) \quad (7)$$

Where 'λ' is a parameter compared to the initially estimated alpha having a large weight while 'γ' is another constant parameter (10−1) and it shows the relative importance in terms of data and smoothness. 'Dg' is a matrix with diagonal values 1 and 0 for known and unknown pixels respectively. 'Γ' is another diagonal matrix with values 0 and 1 for known and unknown pixels respectively.

The confidence value $'\tilde{C}'$ is calculated for each unknown pixel which has been calculated in eq. (8), this value is related to the eq. (6) for the selected foreground 'Fk' and background 'Bk' set for every unknown pixel 'k'.

$$\tilde{C}(F\_(k,)\ B\_(k,)\ \alpha\_(k)) = \exp\left((-(I - \{\alpha\_(k)\ F\_(k) + (1 - \alpha\_k)\ B\_k\}))/(2\sigma\hat{}2)\right) \quad (8)$$

Where 'αk' and foreground and background pair values (Fk, Bk) are obtained in eq. (6) and 'σ' is a constant parameter and set to 0.1 for optimum alpha matte, different values of 'σ' are checked and found optimum result on σ=0.1.

IV. RESULT AND EVOLUTION

The proposed algorithm is tested and implemented on a range of images having different scene intricacies. The test set inclusive ground truth alpha matte is obtained from reference [6].

The proposed method is tested and compared with five state-of-the-art image matting methods:

1. Closed-form Matting (CF) [11],
2. Robust Matting (RB) [16],
3. Learning Based (LB) [37-39],
4. KNN Matting (KNN) [12] and
5. Shared Matting (SM) [18].

A. Qualitative Evaluation

Five images from the given test data set are taken for qualitative measures and evaluation as shown in figure VII [16]. Images img_01, img_04 (doll), and img_02 (trolls) keep a simple background along with a complex foreground, image img_02 contains two connected objects in the foreground also having a large unknown region. These three images contain intricate structures like hairs. Better results were not produced by Closed Form and Learning Based as they included some of the background portions in the foreground. The proposed method produced better results as compared to others. Image img_03 (bear) is slightly simple as there is neither much color variation in the background nor intricate structures. The approach of shared matting and learning-based did not produce better results as they included little of the background portion in the foreground. The approach KNN matting produced results with less error and shared matting misclassified some portion of the caps buckle while the result produced by the proposed technique is found visibly near the Ground Truth. The last image img_05 contains sharp boundaries as shown in the broom caught by the toy other techniques produced misclassified alpha matte while the proposed method produced better.

B. Quantitate Evaluation

The proposed algorithm produced a good result as shown in figure VII. Visual representation is not adequately possible for difference error of estimated alpha with its magnitude in relation to the ground truth so the quantitative comparison is shown.

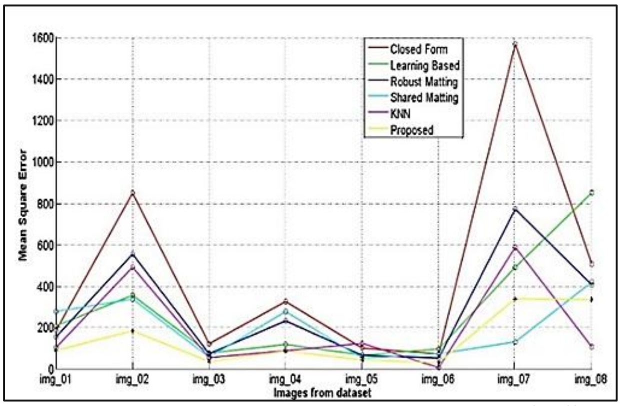

Figure VII: MSE in Alpha Matte against the Ground Truth of Different Algorithms

Table I represents the Mean Squared Error (MSE) of test images against the given ground truth in the dataset and Figure VIII shows a graphical representation of MSE.
The mean square errors are calculated just for the unknown grey region and the obtained alpha value ranges from 0 to 255. The Mean Square Error of eight images is mentioned in Table I.
The MSE is calculated only for unknown pixels as in [16]. Although MSE is not always correlated to the visual matte quality; it still produces a sensible error comparison. The proposed technique produced optimum alpha mattes as shown in figure MSE indicating the MSE of images.

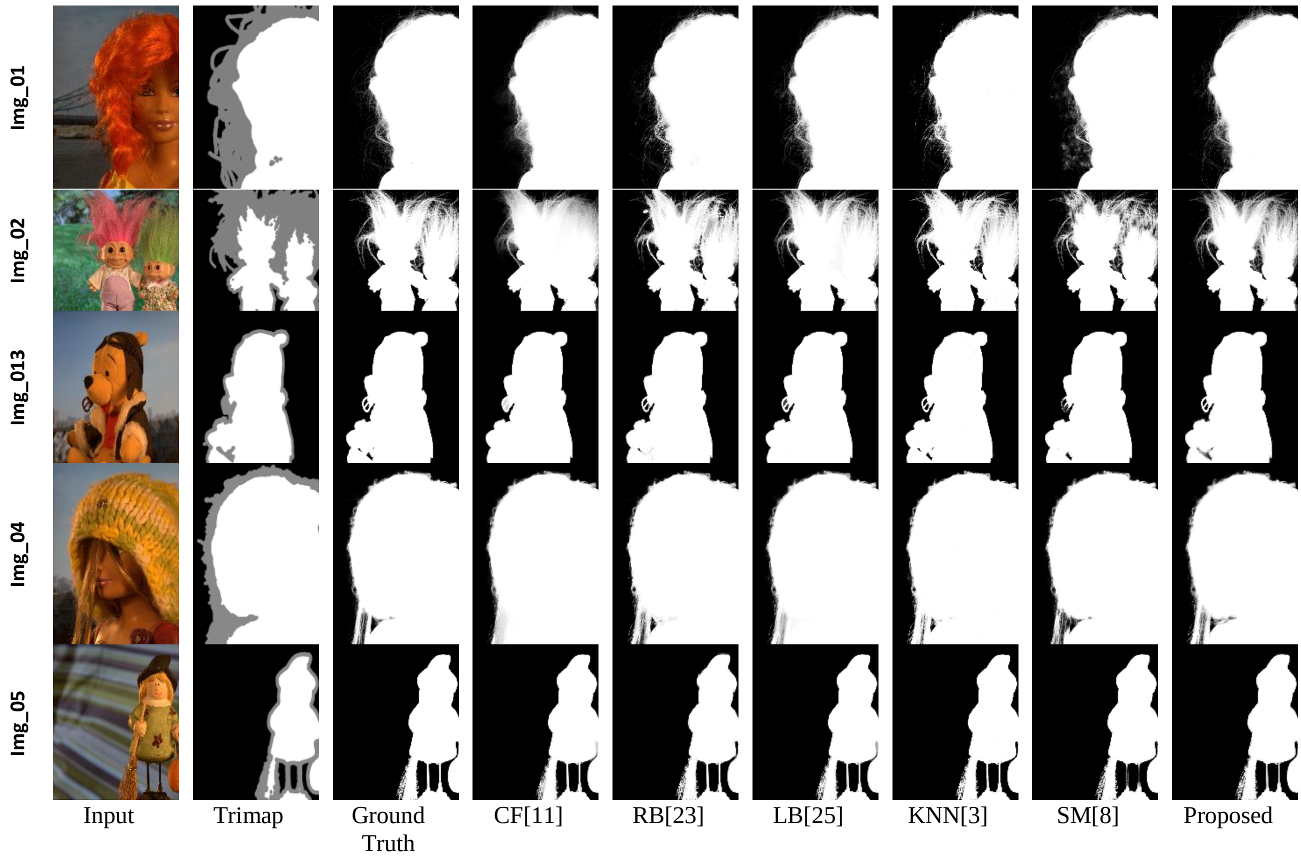

Figure VIII: Estimated Alpha Matte for Visual Comparison with Different Matting Algorithms

Table I: MSE in Alpha against the Ground Truth of Images Taken from Different Algorithms [6]

	CF [11]	RB [23]	LB [25]	SM [8]	KNN [3]	Proposed
Img_01	188	210	157	279	106	**91.9**
Img_02	848	357	555	339	492	**186.3**
Img_03	124	78.7	76.0	59.7	56.9	**38.7**
Img_04	327	120.7	234	279	90.9	**89.2**
Img_05	103	69.8	68.3	54.0	126	**46.7**
Img_06	75.4	98.8	55.1	75.9	**9.42**	33.5
Img_07	1564	492	771	**133**	588	340
Img_08	508	850	407	421	**109**	336

V. CONCLUSION

The superpixel technique in image matting is used as the input of this work. Image matting using superpixel has not been inspected before, as per our knowledge. Superpixel centroid for the finest samples of background and foreground is used in this research. The suggested algorithm is used to acquire enhanced alpha mattes of complex images, as shown in qualitative and quantitative evolution on different and natural images. The efficiency and optimized performance of the proposed method is comparable to the standard image matting algorithm for textured and intricate boundaries. The work would be extended by applying preprocessing of trimap refinement to improve the results. In the future, the proposed work will be extended to a real-time array of images using machine and deep learning algorithms.

Acknowledgment

The authors would like to thank the Department of Computer Science, Federal Urdu University for Arts Science and Technology, Karachi, Pakistan, for their support and assistance throughout this study.

Authors Contributions

The contribution of the authors was as follows: Anam Akbar’s contribution to this study was the concept, technical implementation, project administration, and correspondence. The methodology to conduct this research work was proposed by Aniqa Shirazi along with data collection and supervision. Muhammad Sarim Farooqui facilitated the data compilation and validation. All authors jointly participated in paper writing.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The testing data is available in this paper.

Funding

This research received no external funding.

References

[1] Rubin, M. (2006). *Droidmaker: George Lucas and the digital revolution*. Triad Publishing Company.

[2] Porter, T., & Duff, T. (1984, January). Compositing digital images. In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques* (pp. 253-259).

[3] Li, X., Liu, K., & Dong, Y. (2018). Superpixel-Based Foreground Extraction With Fast Adaptive Trimaps. *IEEE transactions on cybernetics*, *48*(9),2609–2619.

[4] Yang, X., Qiao, Y., Chen, S., He, S., Yin, B., Zhang, Q., ... & Lau, R. W. (2020). Smart scribbles for image matting. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(4), 1-21.

[5] Yao, G. (2017). A Survey on Pre-Processing in Image Matting. *Journal of Computer Science and Technology*, *32*, 122-138.

[6] Rhemann, C., Rother, C., Wang, J., Gelautz, M., Kohli, P., & Rott, P. (2009, June). A perceptually motivated online benchmark for image matting. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 1826-1833). IEEE.

[7] Yao, G. (01 2017). A Survey on Pre-Processing in Image Matting. *Journal of Computer Science and Technology*, *32*, 122–138. doi:10.1007/s11390-017-1709-z

[8] Lin, F.J., & Chuang, J.H. (2018). Alpha Matting Using Robust Color Sampling and Fully Connected Conditional Random Fields. *Multimedia Tools Appl.*, *77*(11), 14327–14342.

[9] Sun, J., Jia, J., Tang, C.K., & Shum, H.Y. (2004). Poisson Matting. In *ACM SIGGRAPH 2004 Papers* (pp. 315–321).

[10] Grady, L., Schiwietz, T., Aharon, S., & Westermann, R. (2005, September). Random walks for interactive alpha-matting. In *Proceedings of VIIP* (Vol. 2005, pp. 423-429).

[11] Levin, A., Lischinski, D., & Weiss, Y. (2006). A Closed-Form Solution to Natural Image Matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*, 228-242.

[12] Chuang, Y. Y., Curless, B., Salesin, D. H., & Szeliski, R. (2001, December). A bayesian approach to digital matting. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (Vol. 2, pp. II-II). IEEE.

[13] Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(5), 603-619.

[14] Ruzon, M.A., & Tomasi, C. (2000). Alpha estimation in natural images. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000* (Cat. No.PR00662), 1, 18-25 vol.1.

[15] Wang, J., & Cohen, M..(2005) An iterative optimization approach for unified image segmentation and matting. . In *ICCV, pages* (pp. 936–943).

[16] Wang, J., & Cohen, M. (2007). Optimized Color Sampling for Robust Matting. In CVPR. *IEEE computer Society*(pp. 1-8).

[17] Rhemann, C., Rother, C., & Gelautz, M. (2008, September). Improving Color Modeling for Alpha Matting. In *BMVC* (Vol. 1, No. 2, p. 3).

[18] Felzenszwalb, P., & Huttenlocher, D. (09 2004). Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, *59*, 167–181.

[19] He, K., Rhemann, C., Rother, C., Tang, X., & Sun, J. (2011, June). A global sampling method for alpha matting. In *CVPR 2011* (pp. 2049-2056). IEEE.

[20] Mishima, Y. (1994). Soft edge chroma-key generation based upon hexoctahedral color space .*U.S. Patent No. 5,355,174*. Washington, DC: U.S. Patent and Trademark Office.

[21] Gastal, E. S. L., & Oliveira, M. M. (2010). Shared Sampling for Real-Time Alpha Matting. *Computer Graphics Forum*, *29*(2), 575–584.

[22] Feng, X., Liang, X., & Zhang, Z. (2016). A cluster sampling method for image matting via sparse coding. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 204-219). Springer International Publishing.

[23] Cho, D., Tai, Y. W., & Kweon, I. S. (2019). Deep Convolutional Neural Network for Natural Image Matting Using Initial Alpha Mattes. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, *28*(3), 1054–1067.

[24] Lu, H., Dai, Y., Shen, C., & Xu, S. (2019). Indices matter: Learning to index for deep image matting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3266-3275).

[25] Qiao, Y., Liu, Y., Yang, X., Zhou, D., Xu, M., Zhang, Q., & Wei, X. (2020). Attention-guided hierarchical structure aggregation for

image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13676-13685).

[26] Levinshtein, A., Stere, A., Kutulakos, K. N., Fleet, D. J., Dickinson, S. J., & Siddiqi, K. (2009). TurboPixels: fast superpixels using geometric flows. *IEEE transactions on pattern analysis and machine intelligence, 31*(12), 2290–2297.

[27] Forte, M., & Pitié, F. (2020). $ F $, $ B $, Alpha Matting. *arXiv preprint arXiv:2003.07711*.

[28] Liu, C., Ding, H., & Jiang, X. (2021). Towards enhancing fine-grained details for image matting. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 385-393).

[29] Kun Wang, Nanning Zheng, & Weixiang Liu (2005). Natural image matting with non-negative matrix factorization. In *Proceedings of the 2005 International Conference on Image Processing, ICIP 2005, Genoa, Italy, September 11-14, 2005* (pp. 1186–1189). IEEE.

[30] Shi, J., & Malik, J. (1997). Normalized cuts and image segmentation. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 731-737.

[31] Hao Wu, Yueli Li, Zhenjiang Miao, Yuqi Wang, Runsheng Zhu, Rongfang Bie, & Rui Lie (2016). A new sampling algorithm for high-quality image matting. *Journal of Visual Communication and Image Representation, 38*, 573-581.

[32] Tang, C. K., Li, D., & Chen, Q. (2012, June). KNN matting. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 869-876). IEEE Computer Society.

[33] Dou Yan, & Xiao Feng, Kong Ling-Fu; Feng. (2009). Improved Knockout Natural Image Matting Algorithm. *Computer Engineering, 35*(14), 212–214.

[34] Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International journal of computer vision, 59*, 167-181.

[35] Vedaldi, A., & Soatto, S. (2008). Quick shift and kernel methods for mode seeking. In *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV 10* (pp. 705-718). Springer Berlin Heidelberg.

[36] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(11), 2274–2282.

[37] Yuanjie Zheng, & C. Kambhamettu (2009). Learning based digital matting. In *2009 IEEE 12th International Conference on Computer Vision*, 889-896.

[38] Assari, S. M., Idrees, H., & Shah, M. (2016). Re-identification of humans in crowds using personal, social and environmental constraints. *arXiv preprint arXiv:1612.02155.*

[39] Cai, S., Zhang, X., Fan, H., Huang, H., Liu, J., Liu, J., ... & Sun, J. (2019). Disentangled image matting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8819-8828).