Non-Parametric patch based video matting

Muhammad Sarim
m.farooqui@surrey.ac.uk
Adrian Hilton
a.hilton@surrey.ac.uk
Jean-Yves Guillemaut
j.guillemaut@surrey.ac.uk

Center of Vision, Speech and Signal Processing
University of Surrey
Guildford GU2 7XH, Surrey,
United Kingdom

Abstract

In computer vision, matting is the process of accurate foreground estimation in images and videos. In this paper we presents a novel patch based approach to video matting relying on non-parametric statistics to represent image variations in appearance. This overcomes the limitation of parametric algorithms which only rely on strong colour correlation between the nearby pixels. Initially we construct a clean background by utilising the foreground object's movement across the background. For a given frame, a trimap is constructed using the background and the last frame's trimap. A patch-based approach is used to estimate the foreground colour for every unknown pixel and finally the alpha matte is extracted. Quantitative evaluation shows that the technique performs better, in terms of the accuracy and the required user interaction, than the current state-of-the-art parametric approaches.

1 Introduction

Matting is a classic problem of image and video processing. Recent advances in digital cameras have increased the interest to develop novel matting techniques in both the image and video domain. Matting is the process of extracting foreground objects while preserving their pixel-wise coverage in the scene. This coverage is referred to as opacity or alpha matte. Once an accurate alpha matte is estimated, a foreground object can be seamlessly composited onto a new background. The matting problem is inherently ill-posed. To make it solvable for an image, skilled user interaction, in the form of a trimap, is often required to aid the definition of foreground and background regions as shown in Fig 1. The task becomes more challenging when image matting is extended to video sequences. Providing a trimap for every frame in a sequence would be too tedious and time consuming. Generally current video matting techniques restrict the requirement for manually defined trimaps to a number of key frames and automatically generate the trimaps for the remaining frames by interpolation. Then the matting algorithm is applied, to individual frames, to estimate the video matte. The matting problem was first formulated by Porter and Duff [12] as linear interpolation of distinct foreground and background images, by using an alpha channel, to

© 2009. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.
BMVC 2009 doi:10.5244/C.23.98

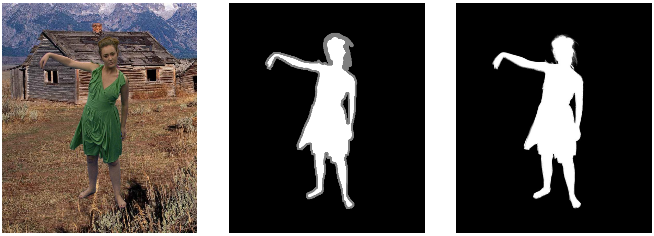


Figure 1: Left: original image, middle: trimap and right: estimated alpha matte.

form a composite image as

$$C_p = \alpha_p F_p + (1-\alpha_p) B_p. \tag{1}$$

This equation is known as the compositing equation where, C_p, F_p and B_p are the composite, foreground and background colours for the pixel p respectively while α_p is their blending proportion. The alpha value ranges from 0 to 1, where $\alpha = 0$ defines the background while $\alpha = 1$ for the foreground. Blended pixels at the foreground boundary have intermediate alpha values. Equation (1) is clearly under-constrained as all the variables on the right hand side are unknown. In a three channel colour space, like RGB, we have only three equations to solve for seven unknowns. In the case of a studio environment, equation (1) can be constrained by using a uniform background, typically blue or green [14]. Enforcing a limitation of no background colour in the foreground provides a trivial solution to the compositing equation. In order to constrain equation (1) for natural images and videos having arbitrary background, user interaction is required to mark some definite background and foreground regions. This definition is referred to as a trimap, where definite foreground and background are represented by *white* and *black* respectively while remaining unknown portions of the image are *gray* as shown in Fig 1. Given a trimap, matting algorithms use local or global image statistics of known regions to compute the alpha values for the unknown region. Existing approaches are usually parametric in the sense that they collect nearby foreground and background colours and fit statistical models to them, such as Gaussian mixture models, to estimate the foreground and background colour for an unknown pixel and finally the α value.

In this paper we present a novel patch based non-parametric approach for video matting. Previously, similar approach have been successfully used to represent local image statistics for inpainting techniques like [6, 7] and view interpolation [8]. Non-parametric patch based sampling provide a strong mechanism to represent local image features, colours and textures which attempts to preserve the spatial information of a natural video sequence. The proposed approach exploits non-parametric statistics for alpha matte estimation in video for both trimap propagation and robust foreground colour estimation. Quantitative comparison to state-of-the-art video matting techniques demonstrates that this approach reduces the error in matte estimation and the amount of manual interaction require to define trimaps.

2 Related work

2.1 Image matting

Sample-based approaches Techniques like [4, 9, 13] fit statistical models to the local known foreground and background pixels which are then used to estimate the foreground and background colour for the unknown pixel and finally compute its alpha value. In a local window, Ruzon and Tomasi [13] modeled the known pixels as a mixture of isotropic Gaussians. These distributions are then used to estimate all the matting variables. Hillman et al. [9], improving on the idea of Ruzon and Tomasi, modeled the known local pixels as anisotropic Gaussian clusters. They used principal components analysis technique to find the major axis of these cluster which are then used for pixel-wise estimation of foreground, background colour and alpha value. Chuang et al. [4] formulated the matting problem in a well known Bayesian framework. Similar to [13] they used Gaussian mixtures to model the known pixels also taking into account the already estimated values of the unknown pixels in a local window. Alpha values are computed by using a maximum a posteriori approach. In Corel Knockout [3] nearby known regions are assumed to be locally smooth. Alpha values are estimated by taking the weighted average of the local known foreground and background pixels. All of these techniques assumed that the known foreground and background regions are locally smooth and strong correlation exists between the nearby known and unknown pixels, raising the requirement for a precise trimap. These algorithms tend to suffer when the local distributions overlap or the unknown region is wide. Some techniques have been proposed which try to generate a good alpha matte for a coarse trimap by using global sampling. Approaches like [2, 16] used a mixture of Gaussians to model foreground and background colour globally. The final alpha mattes are extracted using these global distributions.

Affinity-based approaches Misclassification of colour samples is the main limitation of sample based approaches. To overcome this problem techniques like [10, 15] based on local affinities have been proposed. The affinities are defined in a very small window containing immediate neighbors, where pixel correlation is strong and generally smoothness assumptions hold. Poisson matting [15] assumes that the intensity changes in the foreground and background are locally smooth. The alpha matte is computed by solving the Poisson equation with a matte gradient field. Spatial partial derivatives of the compositing equation (1) approximate the matte gradient field. Local smoothness assumptions allow the Closed-form approach [10] to fit a linear model to the foreground and background colours in a local window, thus defining a quadratic cost function in α. Alpha is then estimated by globally minimising this cost function. Robust matting [17] is a hybrid of local colour sampling and affinity approach. It applies optimised colour sampling technique to the local sparse samples to extract higher confidence sample pairs. It combines affinity similar to [10] with the obtained higher confidence pixels to get the matting energy function which is then minimized to estimate alpha values. This hybrid approach is robust against outliers. The major disadvantage of affinity based techniques is their propagation behavior in α matte estimation. Due to this approach small errors could result in large accumulated errors in the final α matte.

2.2 Video matting

Extracting a foreground object from a single image is a hard problem which becomes even more challenging and difficult for dynamic foreground objects in a video sequences. In general any image matting technique can be used for a video sequence by providing a trimap on per-frame basis. Feeding an algorithm with user defined trimaps for every frame in the sequence is tedious and prohibitively time consuming. Spatio-temporal coherence among the consecutive frames of a sequence can be used to alleviate this difficulty. Different semi-automated techniques have been proposed in existing matting algorithms to reduce the trimap

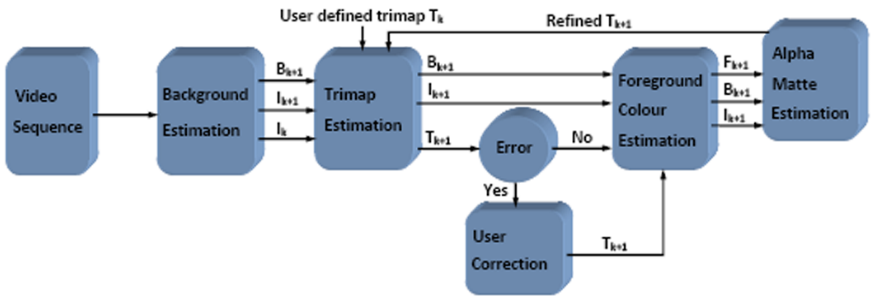


Figure 2: Flow chart for the Non-parametric video matting

construction burden on the user. Often the techniques are split into two steps. In the first step an algorithm interpolates trimap of the intermediate frames using a set of key frames with user defined trimaps. Image matting techniques are then applied to generate alpha mattes for the entire sequence.

Optical flow has been widely used to estimate the inter-frame motion at pixel level in a video sequence. Previously Bayesian video matting [5] successfully utilised optical flow to propagate trimaps from user defined key frames to the rest of the video sequence. Results of optical flow are often erroneous especially for large blurry motions. To ensure the stable propagation of a trimap across a sequence, another initial step is introduced. The step requires the user to provide a "garbage matte" for a sequence that eliminates the foreground object. The remaining background in the sequence is used to construct a mosaic to extract a clean background plate. Optical flow along with the estimated background plate is used to smoothly propagate key frame trimap to the rest of the frames in a sequence. Recently proposed techniques based on rotoscoping [1] and graph cut [11, 18] are also been used as a semi-automated trimap generating system for a video sequence.

3 Non-parametric patch based video matting

Our technique is split into four main steps: (1) constructing a background for every frame of the video sequence, (2) generating a trimap for each frame using the constructed background and the trimap of the previous frame, (3) estimating the foreground colour for every unknown pixel using patch based sampling and (4) generating an alpha matte. The flow chart of the technique is shown in the Fig 2.

3.1 Background construction

To construct a background for planar motion, such as when foreground object is moving across a background, we have used a similar approach to [5]. Background is estimated for every frame in the sequence by utilising optical flow in a very conservative fashion. First a user defines a background region, present near the frame edges, for the few initial frames. A background mosaic is constructed by comparing the defined background region between the successive frames. The background plate for every frame is finally extracted from this mosaic. If the foreground objects in the sequence do not exhibit large motion, an inpainting technique similar to [6] can be adopted to estimate the background for the user defined fore-

I^k I^{k+1} T^k B^{k+1} ${}^1T^{k+1}$ ${}^2T^{k+1}$ ${}^3T^{k+1}$ T^{k+1}

Figure 3: Trimap propagation, red pixels are unlabeled. ${}^1T^{k+1}$ is the result of background comparison, ${}^2T^{k+1}$ is the result of pixel-wise foreground comparison to I^k, ${}^3T^{k+1}$ is the result of foreground patch comparison and T^{k+1} is the final refined trimap.

ground region. In a studio environment, where video matting has a key importance, normally a background image is taken after removing all the foreground objects from the scene. In this paper we have used both, where appropriate, optical flow and studio environment to obtain a clean background plate for every frame. Let us denote the background of the i^{th} frame, I^i, of the original sequence by B^i.

3.2 Trimap propagation

Unlike other techniques, where trimaps for the entire sequence are constructed by initially providing a set of key frame trimaps, our approach automatically constructs a trimap from the previous frame and the background plate and only requires user interaction to define a key frame when automatic propagation fails. After obtaining a trimap, the matting algorithm is applied to estimate the alpha values. The generated alpha matte is then used to refine the trimap prior to its propagation to the next frame. This scheme helps to reduce the number of key frames required and the accumulation of error. Let us denote the trimap of the i^{th} frame, I^i, of the original sequence by T^i. Initially the user defines a fine trimap T^k for the frame I^k, normally the first frame, of the sequence. Let us represent the foreground and the unknown pixels in the trimap, T^k, by the pixel set $(FU)^k$. To propagate the trimap to the next frame, I^{k+1}, the background B^{k+1} is subtracted from the frame I^{k+1}. All the pixels having a difference below a pre-defined distance threshold are marked as definite background pixels in T^{k+1} shown as ${}^1T^{k+1}$ in Fig 3. The remaining pixels are now either definite foreground or unknown denoted by $(FU)^{k+1}$. The Euclidean distance between I^k and I^{k+1} in RGB space for the pixels in $(FU)^k \bigcap (FU)^{k+1}$ is calculated. The trimap value is propagated, from T^k to T^{k+1}, if the pixel-wise Euclidean distance is less than the pre-defined threshold shown as ${}^2T^{k+1}$ in Fig 3.

Normalised sum of square difference (NSSD) is used to associate a trimap label to the pixels still unassigned in T^{k+1}. A square patch, ψ_p, of dimensions n is centred at an unmarked pixel p in the frame I^{k+1}. A patch set ϕ is constructed by localizing square patches, dimensionally consistent to ψ_p, at all the pixels in the set $(FU)^k$ within a spatial radius R to the pixel corresponding to p in the frame I^k. The value of R depends on the inter-frame foreground object motion. If $\phi = \emptyset$, the pixel p is marked as a background pixel otherwise the most similar patch, ϕ_q, to ψ_p in the set ϕ is found by

$$\phi_q = \arg\min_{\phi_i \in \phi} \frac{1}{n^2}\, d\left(\psi_p, \phi_i\right) \tag{2}$$

where, $d\left(\psi_p, \phi_i\right)$ is the sum of square difference between ψ_p and ϕ_i and n^2 is the number of pixels in the patch for normalization. The trimap value T_p^{k+1} of the pixel p is assigned as

$$T_p^{k+1} = \begin{cases} T_q^k & if,\ d\left(I_q^k, I_p^{k+1}\right) \leq \varepsilon \\ unknown & otherwise \end{cases} \tag{3}$$

where, T_q^k is the trimap value of pixel q in the trimap T^k and ε is the pre-defined distance threshold in RGB space. The process is iterated until all the unmarked pixels are assigned a trimap value as it is shown as ${}^3T^{k+1}$ in Fig 3. Assuming that the body of the foreground object is opaque, a final refining step is applied to fill in the unknown holes present within the foreground and background regions. The step is accomplished by applying a connectivity test to the unknown pixel. If the pixel is not 8-connected through the unknown pixels to the foreground boundary, the pixel is given the spatially closest known label. The final trimap, T^{k+1}, is shown in Fig 3.

3.3 Foreground colour estimation

A similar patch based approach to that used in trimap propagation is utilized to estimate the foreground colour for every unknown pixel in the trimap. A square patch ψ_p of dimensions n is centred at an unknown pixel p. A foreground patch set θ is constructed in a similar fashion to the patch set ϕ, as explained in the section 3.2, by using a search window of radius R_f and localising a patch only at the known foreground pixels. The value of R_f depends on the spatial Euclidean distance between the pixel p and the foreground boundary. To find the most similar patch θ_q, the comparison is performed only between the unknown and foreground pixels in ψ_p and the corresponding known foreground pixels of the patch in θ. Let us denote these pixels by p^{uf}. The patch θ_q can be found as

$$\theta_q = \underset{\theta_i \in \theta}{\arg\min}\ \frac{1}{n_{p^{uf}}}\ d_{p^{uf}}\left(\psi_p, \theta_i\right), \tag{4}$$

where, $n_{p^{uf}}$ is the number of pixels p^{uf} which is used for normalization. $d_{p^{uf}}\left(\psi_p, \theta_i\right)$ is the sum of square difference in RGB space between the pixels p^{uf}. The foreground colour, $\tilde{f}$, for the pixel p is approximated as the colour at pixel q. The partial comparison of ψ_p ensures finding similar foreground structure in the known foreground region present in the template ψ_p. To avoid segmentation inaccuracies, which may arise due to the presence of noise in the foreground region, an additional step is introduced for robust foreground colour estimation.

3.3.1 Robust foreground colour estimation

The normalised sum of square difference can be written as

$$\delta_i = \frac{1}{n_{p^{uf}}}\ d_{p^{uf}}\left(\psi_p, \theta_i\right). \tag{5}$$

To precisely estimate the foreground colour for the pixel p, the set of NSSD values δ, is sorted such that $\delta_j < \delta_{j+1}$. We only consider the centre pixel colour of the N most similar patches, written as $\tau = \{\theta_1^c, \theta_2^c, .., \theta_N^c\}$. To remove the effect of noise, the foreground colour $\tilde{f}$ for the pixel p is estimated as the median of τ that is, $\tilde{f} = \mu_{1/2}(\tau)$. The value of N depends on the noise level in the sequence, in this paper we have used $N = 3$. The process is iterated until the foreground colour is approximated for all the unknown pixels.

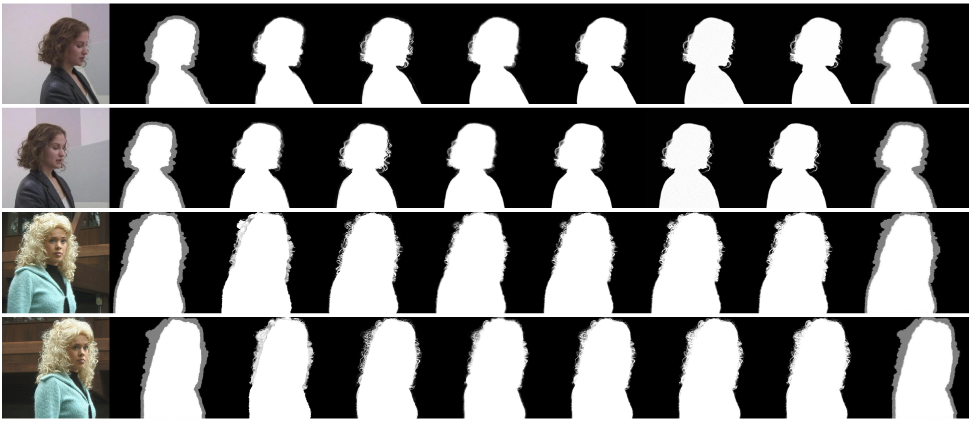


Figure 4: Distinct images 30 frame apart in two different natural video sequences along with their alpha mattes generated by different techniques. The sequence are taken from [5].

3.4 Alpha matte estimation

The final alpha matte is generated by estimating the alpha value for all the pixel in the unknown region of the trimap. The α value for pixel p in the unknown region is computed by rearranging the compositing equation (1) as

$$\alpha_p = \frac{c_p - \tilde{b_p}}{\tilde{f_p} - \tilde{b_p}}. \quad (6)$$

Where, c_p and $\tilde{f_p}$ are the composite and approximated foreground colour respectively while $\tilde{b_p}$ is the estimated background colour from the background plate extracted in the section 3.1. Once an alpha matte is computed, the foreground object can be seamlessly composited onto a new background.

4 Results and evaluation

We present a detailed comparison of the proposed technique with other established matting algorithms. We have used two natural video sequences used in previous matting papers [5, 16] for the qualitative comparison while three composite sequences for quantitative evaluation. The composite videos are captured in a studio environment with a uniform blue background in order to provide precise ground truth. The ground truth alpha mattes are generated by applying Closed form [10] matting algorithm to user defined precise trimaps. The ground truths are used to form composite video sequences according to equation (1). For the sake of fair comparison we have utilised an approach similar to [5] to generate trimaps for the entire sequence by providing key frame after every 10 frames in the presented videos. We have used Hillman [9], Poisson [15], Closed-form [10] and Robust matting [17] algorithms for comparison. Our matting technique is applied in two different ways for analysis: (1) using the trimaps generated for other approaches as explained above, refered to as *Non-para 1* and (2) implementation of our complete algorithm including the trimaps refered to as *Non-para 2*.

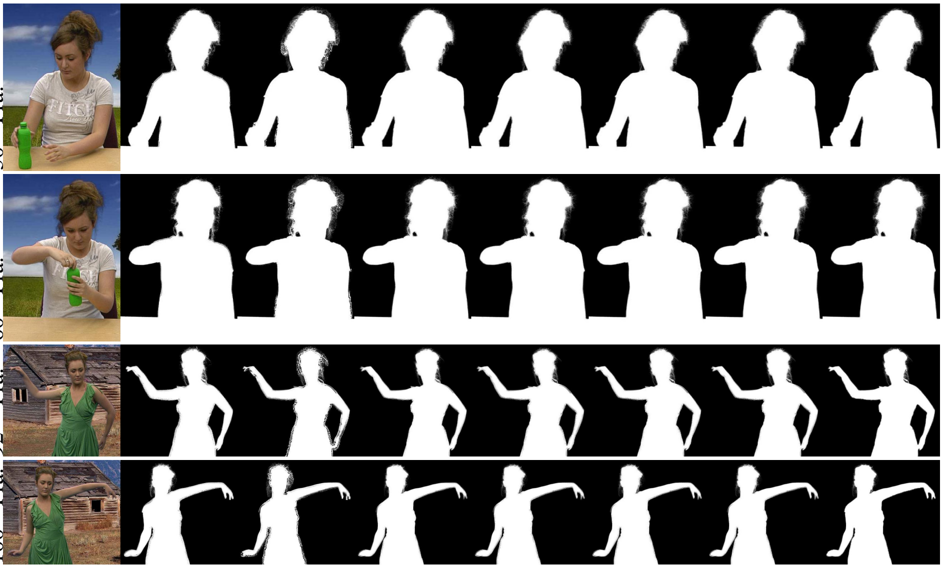


Figure 5: Frames from two of the three different composite video sequences along with their alpha matte generated by different techniques and ground truths.

4.1 Qualitative evaluation

Fig 4 shows images from two different natural video sequences and their alpha mattes computed by different techniques. For the first sequence all the techniques except Poisson produced acceptable results mainly because of the simple background and distinct foreground colour distribution. The global optimization of Poisson matting generated the segmentation in the blended region. Hillman et al.'s approach produced unacceptable blurred matte for the second video sequence due to the presence of large unknown regions and local foreground and background distribution overlap in colour space. Our technique along with the Closed-form and Robust matting algorithms, generated mattes which are visibly smooth and perceptually indistinguishable, with a reduced interactive requirement in the case of *Non-para 2*.

	Office	Dance	Walk	*Av. rank*
Hillman	$59.81^{6} : 05.39^{5}$	$49.18^{5} : 09.59^{1}$	$39.53^{5} : 13.67^{2}$	5.33 : 2.66
Poisson	$53.35^{5} : 07.33^{6}$	$97.60^{6} : 17.13^{6}$	$79.57^{6} : 22.39^{6}$	5.66 : 6.00
Clo-fo	$08.60^{3} : 04.80^{4}$	$30.61^{3} : 13.32^{5}$	$20.73^{3} : 14.36^{3}$	3.00 : 4.00
Robust	$09.88^{4} : 02.86^{1}$	$38.19^{4} : 12.13^{3}$	$28.55^{4} : 15.09^{4}$	4.00 : 2.66
N-para-1	$03.62^{1} : 03.98^{2}$	$28.35^{1} : 11.03^{2}$	$18.72^{2} : 13.23^{1}$	1.33 : 1.66
N-para-2	$03.89^{2} : 04.56^{3}$	$29.20^{2} : 12.37^{4}$	$18.53^{1} : 16.65^{5}$	1.66 : 4.00

Accuracy and robustness table in the format $RMSE_{min}^{rank} : \Delta RMSE^{rank}$

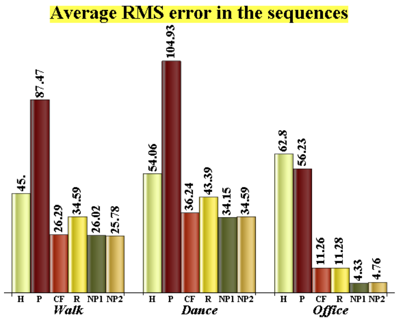
(a)

	Fra.	Key oth.	Key NP2.
Natu-1	145	15	4
Natu-2	91	10	3
Office	125	13	6
Dance	125	13	8
Walk	101	11	4

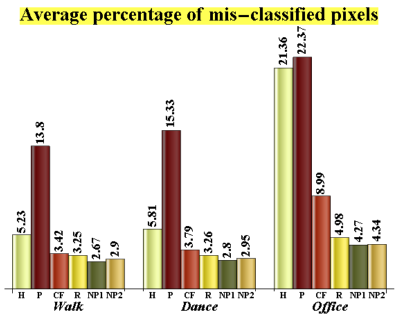
Number of key frames used

(b)

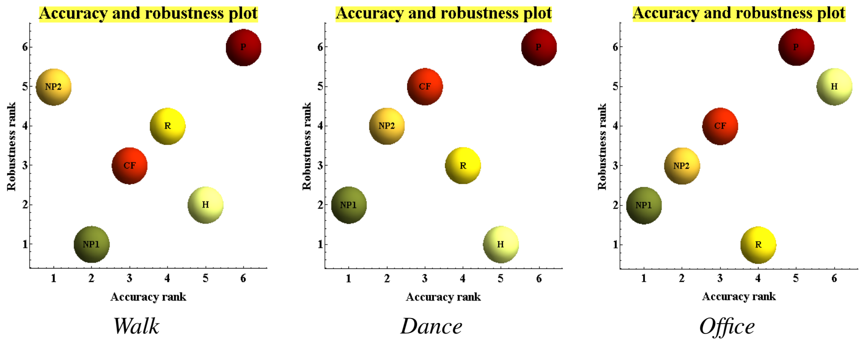
Table 1: (a) Accuracy and robustness rank table, (b) number of key frames used in implementation of other techniques and Non-para 2.



(a)



(b)



(c) Accuracy and robustness rank plot, smaller values represent better performance.

Figure 6: Techniques are refered to by their initials. Alpha is scaled to $[0-255]$ for RMSE.

4.2 Quantitative evaluation

For the quantitative comparison we have used three composite video sequences as can be seen in Fig 5, having length between 100 and 125 frames as shown in table 1(b), generated as explained in section 4. We have used two error measurements: (1) Root mean square error, RMSE and (2) the percentage of pixels misclassified either as foreground or background. Fig 6(a,b) shows the RMSE and the percentage of misclassified pixels for the three sequences produced by different techniques respectively. The RMSE is also used to evaluate the accuracy and the robustness of the techniques. The minimum RMSE, $RMSE_{min}$, in a given sequence represents its accuracy while the difference $\Delta RMSE = RMSE_{max} - RMSE_{min}$ represents the robustness of an algorithm. Table 1 shows the accuracy and robustness rank of different techniques and the number of key frames used in the complete implementation of our algorithm and other techniques. Although it is difficult to visually distinguish, the result of Closed-form, Robust and Non-parametric algorithms, the charts show that our approach has produced results that have lower error. In the sequence where the foreground motion is large, it is not surprising that $Non-para1$ produced slightly better results, than $Non-para2$, because of the higher number of key frames available. An advantage of $Non-para2$ compared to $Non-para1$ is that it defines key frame adaptively. This considerably reduces the number of required key frames compared to techniques using regularly sampled key frames, as can be seen in table 1(b), while producing results which are qualitatively and quantitatively similar. The rank plots for the used sequences are shown in the Fig 6(c). Our algorithm manages to produce results which are perceptually similar to the ground truth and quantitatively more precise than other state-of-the-art techniques.

5 Conclusion

A novel patch based non-parametric video matting technique is presented. We have used optical flow in a conservative manner to construct the background and to propagate the trimap. Sampling local patches rather than fitting statistical models, to the defined image regions, helps our technique to preserve the spatial information of the natural scenes not only in colour but also in image structure and texture space. The patch based approach diminishes the requirement for local smoothness and correlation assumptions made by other state-of-the-art matting techniques. A detailed evaluation shows that our approach has a clear advantage over parametric techniques both in terms of foreground, background colour estimation and user interaction required even for a large foreground motion. Affinity based techniques tend to produce comparable results but they suffer from accumulation of error as they estimate alpha values in a propagation manner. Future work in non-parametric matting will focus on developing a more robust matching criteria to deal with the moving background objects. Smoothness constrains will also be incorporated to further optimize the perceptual quality of the alpha matte.

References

[1] A. Agarwala, A. Hertzmann, D. H. Salesin, and S. M. Seitz. Keyframe-based tracking for rotoscoping and animation. *ACM Trans. Graph.*, 23(3):584–591, 2004. ISSN 0730-0301. http://doi.acm.org/10.1145/1015706.1015764.

[2] X. Bai and G. Sapiro. Geodesic matting: A framework for fast interactive image and video segmentation and matting. *Int. J. Comput. Vision*, 82(2):113–132, 2009. ISSN 0920-5691. http://dx.doi.org/10.1007/s11263-008-0191-z.

[3] A. Berman, A. Dadourian, and P. Vlahos. Method of removing from an image the background surrounding a selected object. U.S. Patent 6,134,346, 2000.

[4] Y. Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *Proceedings of IEEE CVPR '01*, volume 2, pages 264–271, December 2001.

[5] Y. Y. Chuang, A. Agarwala, B. Curless, D. Salesin, and R. Szeliski. Video matting of complex scenes. In *Proceedings of ACM SIGGRAPH*, pages 243–248, 2002.

[6] A. Criminisi, P. Pérez, and K. Toyama. Object removal by exemplar-based inpainting. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:721–728, 2003.

[7] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *IEEE International conference on computer vision*, pages 1033–1038, 1999.

[8] A. Fitzgibbon, Y. Wexler, and A. Zisserman. Image based redering using image based priors. In *International conference on computer vision ICCV*, pages 1176–1184, 2003.

[9] P. Hillman, J. Hannah, and D. Renshaw. Alpha channel estimation in high resolution images and image sequences. In *IEEE CVPR*, pages 1063–1068, 2001.

[10] A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:61–68, 2006. ISSN 1063-6919. http://doi.ieeecomputersociety.org/10.1109/CVPR.2006.18.

[11] Y. Li, J. Sun, and H. Y. Shum. Video object cut and paste. *ACM Trans. Graph.*, 24(3):595–600, 2005. ISSN 0730-0301. http://doi.acm.org/10.1145/1073204.1073234.

[12] T. Porter and T. Duff. Compositing digital images. In *ACM SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 253–259, 1984.

[13] M. A. Ruzon and C. Tomasi. Alpha estimation in natural images. In *CVPR*, pages 18–25, June 2000.

[14] A. R. Smith and J. F. Blinn. Blue screen matting. In *ACM SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 259–268, 1996.

[15] J. Sun, J. Jia, C.K. Tang, and H. Y. Shum. Poisson matting. *ACM Transactions on Graphics*, 23 (3):315–321, 2004.

[16] J. Wang and M. F. Cohen. An iterative optimization approach for unified image segmentation and matting. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 936–943, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2334-X-02. http://dx.doi.org/10.1109/ICCV.2005.37.

[17] J. Wang and M. F. Cohen. Optimized color sampling for robust matting. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8, 2007.

[18] J. Wang, P. Bhat, R. A. Colburn, M. Agrawala, and M. F. Cohen. Interactive video cutout. *ACM Trans. Graph.*, 24(3):585–594, 2005. ISSN 0730-0301. http://doi.acm.org/10.1145/1073204.1073233.