Sheraton Surabaya Hotel & Towers, October 17th - 19th, 2019
IJCST
International Joint Conference on Science and Technology
Chapter ICST (International Conference on Science and Technology)

Soft Segmentation Using Image Registration Based on Nonlocal Matting

Meidya Koeshardianto

Informatics Engineering, Trunojoyo University of Madura

meidya@trunojoyo.ac.id

Abstract. Soft segmentation is used for extracting foreground and background on image or video sequence. In this paper, we proposed a new method of soft extraction from an image or video using image registration based on nonlocal matting. One of the disadvantages in nonlocal matting is needed to obtain constraint manually. The constraint is used to identify foreground and background area on separate image region. Based on the registration method, we try to reduce the cost to determine it. First, a constraint is obtained with identify key points detection and descriptor between image source and template. Image template was provided before extracting. Then soft segmentation, nonlocal matting method, is used to extract region. In our experiment result, we show that our proposed framework has significant performance improvement over the state-of-the-art.

1. Introduction

Separate foreground and background in movie maker, especially in visual effect animation is a difficult task. Extracting foreground has been used to obtain the object; therefore, it could be composited with the background and taken in the real condition.

The separation process between the foreground and background can be done using some techniques, such as image segmentation. Nevertheless, image segmentation [1],[2] results still cannot separate objects accurately. Image matting has been proposed as a solution to solve this problem of object separation.

Image matting is one of the methods that aim to do soft segmentation. The earlies work in image matting was done by assuming the green color as the background [3]. However, this approach is rarely used in the recent image matting research because it has a limitation which cannot retrieve image foreground naturally.

Natural image matting was solved by proposing two thoughtful approaches that are supervised and unsupervised matting. The supervised methods need defined foreground and background constraints before applying matting. These methods have advantages that foreground and background can be determined accurately, but it requires effort for manual labeling using trimap or scribbles.

Unsupervised matting reduces complexity by eliminating effort for labeling the constraints. However, in some cases, the results of the matting process are incompatible with user intention so that the extraction process does not match with their desire.

Considering all the problem above, the main contribution of this paper implies a supervised approach with image template guidance. A constraint is determined by the similarity feature between image and template. Next, nonlocal matting is obtained the alpha matte that is used to extract foreground and background.

The rest of this paper is organized as follows. We introduce our proposed methods and our contribution in section I. We review current research in image matting methods in section II. Section III explore our proposed method on image registration such as constraint building. Section IV explanation detailed of the proposed method. Section V shows the experimental result. We conclude the paper in section VI.

2. Current Research in Image Matting

Formally, matting methods were described by Potter [4] that take as input I is assumed to be a composite of a foreground F and a background B as in equation 1. It is a color of the i^{th} pixel that assumed to be a linear combination of the corresponding colors.

$$I_p = \alpha_p F_p + (1 - \alpha_p B_p) \quad (1)$$

where $p = (x, y)$ and α_p is the pixel's foreground opacity.

Ruzon et.al [5] have modeled the foreground and background as a mixture of un-oriented Gaussians. Furthermore, a Bayesian matting method [6], the foreground and background distributions are modeled using a mixture of oriented Gaussians and a maximum-likelihood criterion, is used to estimate the final matte.

In affinity-based matting methods solve alpha values by defining various affinities between nearby pixels. The Poisson matting method solves the matte from the relative gradient field estimated from a given image [7]. The image can be modified by treating the gradient interactively or automatically. Then the intensity changes in both foreground and background locally smooth.

Then, L.Grady et.al [8] which computes the alpha values based on the affinity between neighboring pixels is calculated with the measurement of the color distances in the RGB channels by using a Local preserving projection. Furthermore, X. Bai et.al [9] approach measures the weighted geodesic distance to estimate alpha values that a random walker will travel from its origin to reach the foreground.

A recent affinity-based method which offers both trimap and scribble based matting with Matting Laplacian matrix has done by [10], [11]. Scribbles play an important role in obtaining a quality matte. In [10] need more input than [11] with the better the solution just despite a pixel on the object. Manual labeling required them for determining scribbles of foreground and background. We used this method for determining alpha matte and will be discussed further in the following section.

Zang et.al [12] identifies false detected shadows by detecting moving feature and corrects classifications errors by means of image matting operated on candidate shadow regions. Then [13] use template matching techniques to recognize the object using correlation and phase angle methods. Moreover, [14] using SIFT and SURF matching and image matting to detect fake shadow. According to this, need more input for determining scribbles, we propose to construct a system for extracting object using automatic scribbling based on feature matching in nonlocal matting.

3. Image Registration as Constraint Building

Registration method is used to find feature matching between two images, image template and source images, at different times and viewpoints. Image registration used for building constraint as a scribbles that are created automatically by adjusting these steps: key point's detection, key point's descriptor and image registration, as shown in figure 1.

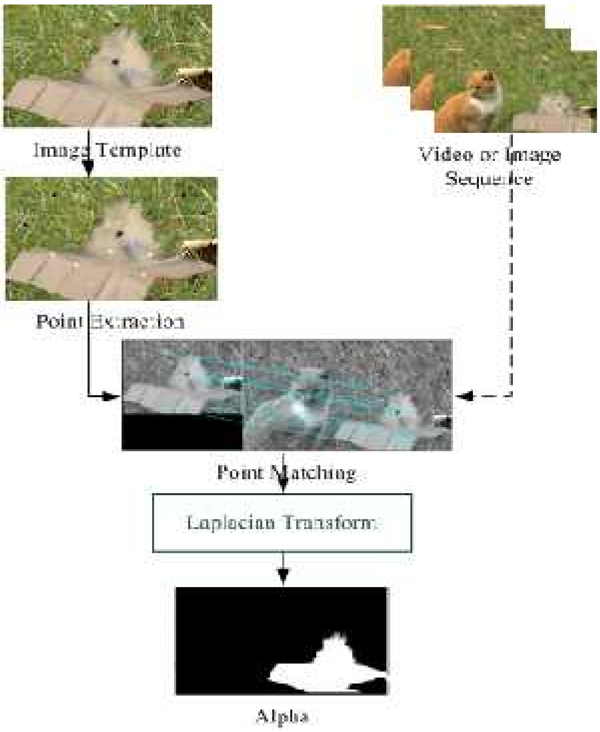

Figure 1. Architecture System

3.1. Key Point's Detection

To obtain key points, DoG (Difference of Gaussian) function $D(x, y, \sigma)$ is used to achieve scale invariance which is the difference of smoothed $G(x, y, \sigma)$ to do convolution on an image [15].

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} * I_p \qquad (2)$$

We determine variable σ to obtain different scale images. We use these simulation results on experiment and discussion section. We used varying the number between 0.2 of 0.9 to examine the effect of scales per octave at which the image function. Then, subtract it to the images which are adjacent to get a DoG pyramid that can write:

$$D(p, \sigma) = G(p, k\ \ ) - G(p, \sigma) \qquad (3)$$

where k denotes a scale coefficient of an adjacent scale-space factor.

3.2. Key point's Descriptor

Next, to achieve key point's descriptor, we compute the gradient strength (4) of every pixel in the scale of each key point using the following expressions:

$$m_p = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \qquad (4)$$

and the direction for each pixel follows as:

$$\theta(x, y) = \tan^{-1} \frac{f_y(x,y)}{f_x(x,y)} \qquad (5)$$

where

$$f_x(x,y) = G(x+1,y) - G(x+1,y)$$

$$f_y(x,y) = G(x,y+1) - G(x,y-1)$$

Key point's descriptor is determined by choosing histogram for every neighborhood and uses the summations as the gradient strengths that have the maximum value of gradient strength. After the key point as a center has chosen an adjacent 16×16 region, then divides this region into 4×4 sub-regions and sums the gradient strength in each sub-region.

3.3. Image Registration

Image registration is done by summing of the gradient strength, feature database is created by extracting key points from model image which is taken from different angles. Then by (6), matching process key points between the feature on image template and images on frames is performed.

$$n' = a \qquad \sqrt{\sum_{P}^{1} (s_p^m - w_p^m)} \qquad (6)$$

The key point n' which makes Euclid distance minimum is obtained so we have determined foreground models for constraint for soft segmentation.

4. Soft Segmentation

In this paper, soft segmentation is used to extract foreground and background object by compute α with *Matting Laplacian*. Nonlocal matting algorithm proposed by P. Lee [11], is used to extract it state that interpretation Laplacian as a graph $g = (v, \epsilon)$ where W_i represent the weight of edge $(i,j) \in \epsilon$.

$$W_i^L = \sum_{k|i,j \in s(k)} \left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\lambda + \sigma_k^2}\right) \qquad (7)$$

where μ_k and σ_k^2 are mean and variance of patch k. Then recall in [11] the following :

$$\epsilon[X_i] \approx \sum_j X_j k_i \frac{1}{D_i} \qquad (8)$$

where

$$k_i = \exp\left(-\frac{1}{h_1^2}\|X_i - X_j\|_g^2 - \frac{1}{h_2^2} d_i^2\right) \qquad (9)$$

and

$$D_i = \sum_j k_i \qquad (10)$$

where X_i is a feature vector that computed by the information around pixel i, and d_i is distance between pixels i and j, $|k|_g$ is a norm weighted by a center-weighted Gaussian, $h_1$and h_2 are some constants found empirically.

Soft segmentation is obtained alpha matte, as in equation (1), component for extracting and compositing object. With equation (8), we obtain the alpha value as:

$$\epsilon[\alpha_i] = \sum_j \alpha_j k_i \frac{1}{D_i} \qquad (11)$$

or we can write:

$$D_i \alpha_i \approx k(i,.)^T \alpha \qquad (12)$$

With the derivation $D = W$, where W as in (7) then $(D - W)\alpha \approx 0$ or $\alpha^T L \approx 0$ so then $L = (D - A)^T (D - A)$ is called clustering Laplacian. The quadratic form $\alpha^T L$ has the particularly nice form

$$\sum_{i,j\in e} W_i (\alpha_i - \alpha_j)^2$$

which is a measure of smoothness along the edges of g. It is directly obvious from L that the vector of all ones 1 is an eigenvector of L having eigenvalue 0.

If there is a subset of pixels that exactly cluster in the graph implied by Laplacian L then the value of the objective function $q(\alpha)$ is minimized if the rest of the pixels in the cluster are labeled α. The reason is that, if we have K clusters, $q(\alpha)$ can be decomposed as:

$$q(\alpha) = \sum_{k=1}^{K} \sum_{i,j\in e_k} W_i (\alpha_i - \alpha_j)^2 \qquad (13)$$

where $e_k | k = 1,2,..,K$ is partition of e.Then labelling the cluster as minimize $q(\alpha)$.

The match key points are used as foreground constraint and background constraint is created from subtraction number of cluster with foreground. From (13) the value background is given as$B_s(k) = e_s(k) - F_s(k)$, where e_s as number of clusters, B_s are background clusters, and F_s are foreground clusters.

It is means that we can construct the affinities so that the graph clusters accurately on foreground and background objects, with only needs to constrain a single pixel in each cluster. It is suitable for using the result image key points to label a single pixel in each of the clusters of L.

5. Experiment Result and Discussion

In this experiment, performance algorithm of all the result is reported on a PC with Intel(R) Core(TM) i5-8250U@1,6 GHz and 8 GB memory, NVIDIA GeForce 940MX (4GB) graphics card.

The proposed algorithm is construct soft segmentation based on image template. The feature SIFT key points are used as foreground scribbles. The other sparse is created from occupied by key points as background scribbles as shown in figure 2.

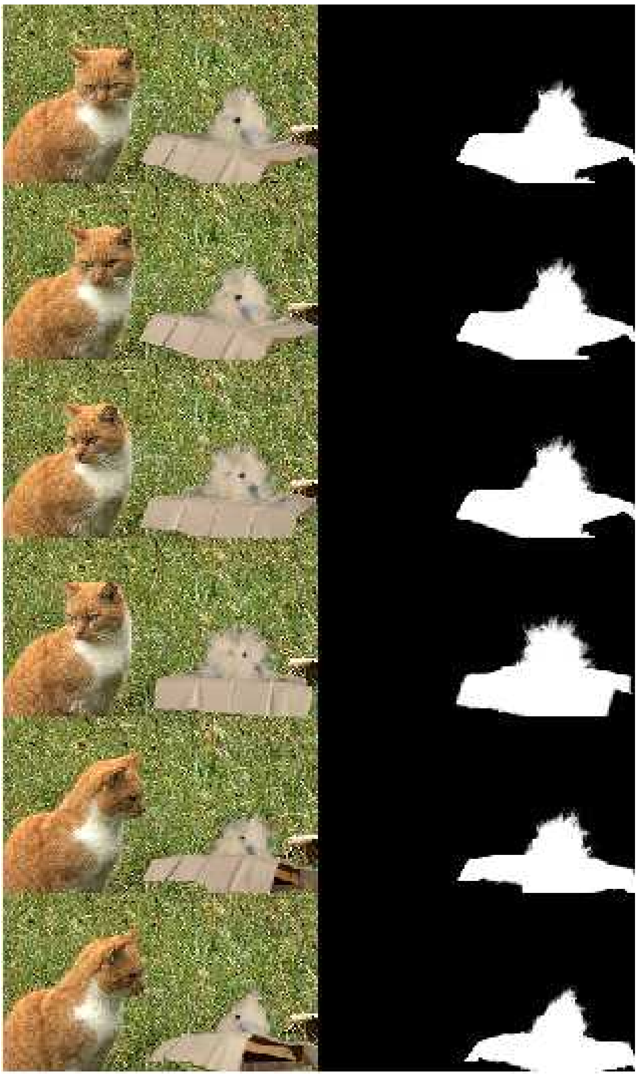

Figure 2. Alpha matte from image sequence. Left images are source frames and the right is alpha matte.

6. Conclusion

The key point of this work mainly focuses on using SIFT as feature matching to provide automatic scribbling for soft segmentation. Our simulations or experiment show that this proposed method can be implemented on nonlocal matting by looking on our result. According from our sample video, it shows that objects can be extracted by most satisfactory parameter using $\theta = 0{,}05$. This experiment also shows that θ is depend on different case in natural image and video matting as an input. Future work includes investigating updated or replaced image template periodically is needed on actual implementation. The shifting image template should be performed cause of changes feature on the next frames in video sequences. So that, the image template feature will be still recognized on the frame.

References

[1] N. M. Saad, N. S. M. Noor, A. R. Abdullah, S. Muda, A. F. Muda, and H. Musa, “Segmentation and classification analysis techniques for stroke based on diffusion weighted images,” *IAENG Int. J. Comput. Sci.*, vol. 44, no. 3, 2017.

[2] M. Beneˇ, “Segmentation of MRI Images by Adaptive,” vol. i, no. July, 2015.

[3] A. R. Smith and J. F. Blinn, “Blue screen matting,” *SIGGRAPH ’96 Proc. 23rd Annu. Conf.*

Comput. Graph. Interact. Tech., pp. 259–268, 1996.
[4] T. Porter and T. Duff, "Compositing digital images," *ACM SIGGRAPH Comput. Graph.*, vol. 18, no. 3, pp. 253–259, 1984.
[5] M. A. Ruzon and C. Tomasi, "Alpha Estimation in Natural Images," *Proc. CVPR*, vol. 1, no. June, pp. 18–25, 2000.
[6] Chuang, Curless, Salesin, and Szeliski, "A Bayesian approach to digital matting," *Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR 2001*, vol. 2, no. 3, p. II-264-II-271, 2001.
[7] J. Sun, J. Jia, C. Tang, and H. Shum, "Poisson matting," *ACM Trans. Graph.*, vol. 23, no. 3, p. 315, 2004.
[8] L. Grady, T. Schiwietz, and S. Aharon, "Random walks for interactive alpha-matting," *IASTED Int. Conf. Vis. Imaging Image Process.*, pp. 423–429, 2005.
[9] X. Bai and G. Sapiro, "Geodesic matting: A framework for fast interactive image and video segmentation and matting," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 113–132, 2009.
[10] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *Pami*, vol. 30, no. 2, pp. 228–42, 2008.
[11] P. Lee and Y. Wu, "Nonlocal matting," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2193–2200, 2011.
[12] H. wang Liang Zang, Xiaomei He, "Shadow Verification Based on Feature Matching and Image Matting," *Igarss 2014*, no. 1, pp. 1–5, 2014.
[13] K. Ahuja and P. Tuli, "Object Recognition by Template Matching Using Correlations and Phase Angle Method," *Int. J. Adv. Res. …*, vol. 2, no. 3, pp. 1368–1373, 2013.
[14] L. Zhang, H. Wang, T. Deng, and X. He, "Improving integrality of detected moving objects based on image matting," *Chinese J. Electron.*, vol. 23, no. 4, pp. 742–746, 2014.
[15] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *Int'l J. Comput. Vis.*, vol. 60, pp. 91–11020042, 2004.