UNSUPERVISED SEGMENTATION OF LOW DEPTH OF FIELD IMAGES BASED ON L_0 REGULARIZED MATTING MODEL

Yibo Chen, Tianle Zhao, Wai-Kuen Cham

The Chinese University of Hong Kong

ABSTRACT

An effective unsupervised segmentation method is proposed to extract the object of interest from low depth of field images based on a novel L_0 regularized matting model. First of all, a multi-scale reblurring model is utilized, together with guided filter and morphological filter, to generate a trimap that roughly labels the focused and defocused regions. Then, an L_0 regularized matting model is proposed to obtain the accurate segmentation of the object of interest. Experimental results demonstrate that the proposed method achieves state-of-the-art performance for unsupervised segmentation under various situations, and is robust to noise.

Index Terms— Image Segmentation, Low Depth of Field, Object of Interest, Matting, L_0 norm

1. INTRODUCTION

In order to emphasize an object of interest (OOI) within an image, photographers often use an optical system with low depth of field (DOF) in practice, which is very popular nowadays. In images captured by this kind of systems, only the OOI parts are well focused while the other parts like background are blurred, which is also called defocused. This will help viewers pay more attention to the OOI and understand the theme of image quickly. Therefore, an effective unsupervised segmentation method that can extract the OOI in a low DOF image will have broad application prospects, such as adaptive image compression, fast indexing and retrieval from image and video databases, and 3D object reconstruction [1, 2, 3].

Segmentation of images with low DOF has attracted more interest in past years and a lot of segmentation methods have been proposed in literatures. Tsai and Wang proposed the first approach based on edge properties [4]. It measures the amount of defocus at each edge pixel and determines the boundary of OOI with an edge-linking algorithm. This method has high accuracy in segmenting objects with clear boundaries, but it fails to tackle complex edges. Therefore many region-based algorithms were proposed. In [5], low DOF images were firstly transformed into wavelet domain and statistical properties of high frequency coefficients were used to detect the focused regions. The final segmentation results were obtained by a multi-scale method using blocks of various sizes. However, this method concentrates too much on high frequency components: smooth regions in the OOI may also be removed and unwanted holes will appear. Graf [6] proposed a parameterless image segmentation method which contains five complicated stages. The complicated procedures provide more restrictions to ensure good segmentation results, but may also increase the risk of information lost. Li [7] proposed a reblurring model to generate a focused saliency map. Then he refined the saliency map with a bilateral filter, and extracted the OOI by applying seeded region growing and merging. Mei [8] further improved this method by using a multi-scale reblurring model and an active contour model based on hybrid energy. Mei's method achieves the state-of-the-art performance, however it has high computational requirement and still has difficulties in handling shadows or smooth regions in the OOI.

In this paper, a novel robust method based on an L_0 regularized matting model is proposed for automatic segmentation of low DOF images. Experimental results demonstrate that the proposed method outperforms current state-of-the-art methods under various situations. The main contributions of the proposed method are highlighted as follows.

(i) The introduction of matting terms in the proposed model better exploits the correlation among pixels, gives a good spreading effect in smooth regions, and finds the accurate boundary of focused regions and defocused regions.

(ii) The introduction of L_0 norm gives a strong restriction at the boundary and reduces sawtooth artifacts. Besides, L_0 norm also helps improve the robustness to noise.

The rest of this paper is organized as follows. Section 2 explains the proposed method in detail. Section 3 gives the solution of the proposed optimization problem and experimental results are shown in Section 4. Section 5 draws some concluding remarks.

2. PROPOSED ALGORITHM

In this section, the proposed method for low DOF image segmentation, which consists of two steps, is presented in detail. In the first step, a focused saliency map is generated by a multi-scale reblurring model. Guided filter and morphological filter are then applied to generate a trimap which separates

the saliency map into focused regions, defocused regions and unknown regions. In the second step, an L_0 regularized matting model is developed to segment the OOI accurately.

2.1. Focused Saliency Map and Trimap Generation

Unsupervised segmentation means separating the OOI apart from background without any human help. To achieve this goal, a method that can automatically detect the focused regions is needed.

Let I denotes the original image, which combines a focused image f and a defocused image b. The original image can be modeled by

$$I(x,y) = \alpha(x,y)f(x,y) + (1-\alpha(x,y))b(x,y), \quad (1)$$

where (x,y) denotes the pixel location, α is an alpha matte considered as a binary function, and is defined as

$$\alpha(x,y) = \begin{cases} 1, & (x,y) \in \Omega_f \\ 0, & (x,y) \in \Omega_b \end{cases}, \quad (2)$$

where Ω_f and Ω_b denote focused regions and defocused regions respectively. It has been proved that the reblurring model proposed by Li [7] and improved by Mei [8] provides reliable information for detecting focused regions. The focused saliency map κ is then generated from a low DOF image as:

$$\kappa(x,y) = |I(x,y) - I_r(x,y)|, \quad (3)$$

where I_r is a reblurred version of I based on a multi-scale reblurring model in [8].

Since focused saliency map is usually sparse and isolated points may influence the segmentation results a lot, guided filter proposed by He [9] is utilized to smooth the saliency map while preserving edge properties according to the original image I.

$$\kappa_g = G_{r,\xi}(\kappa, I), \quad (4)$$

where r and ξ are the parameters of the guided filter, κ_g is the resulting saliency map after the guided filter $G_{r,\xi}$. Then the saliency map is separated into focused regions Ω_f, defocused regions Ω_b and unknown regions Ω_u, by measuring the filtered saliency value of each pixel

$$(x,y) \in \begin{cases} \Omega_f, & \text{if } \kappa_g(x,y) \geq t_1 \\ \Omega_b, & \text{if } \kappa_g(x,y) \leq t_2 \\ \Omega_u, & \text{otherwise} \end{cases}. \quad (5)$$

In our experiments, t_1 is set to be 0.3 and t_2 is set to be 0.1. Since the saliency map may not be accurate at the boundary of focused regions and defocused regions, especially for the identification of focused regions, morphological filters are applied to reduce the influence of inaccurate saliency values.

$$\tilde{\Omega}_f = \mathcal{M}_e(\mathcal{M}_d(\Omega_f, r_d), r_e), \quad (6)$$

where $\mathcal{M}_e$ and $\mathcal{M}_d$ are the morphological erosion and dilation operators respectively. r_e is the size of erosion kernel and r_d is the size of dilation kernel. After the morphological operations, the three regions are refined as

$$\begin{array}{rcl} \Omega_F &=& \tilde{\Omega}_f - \tilde{\Omega}_f \bigcap \Omega_b \\ \Omega_B &=& \Omega_b - \tilde{\Omega}_f \bigcap \Omega_b \\ \Omega_U &=& \Omega - (\Omega_F \bigcup \Omega_B) \end{array}, \quad (7)$$

where Ω is the whole image. Then a trimap α_s is generated, where focused pixels in Ω_F are labelled as 1, defocused pixels in Ω_B are labelled as 0, and the other pixels in Ω_U remain unlabelled.

2.2. L_0 Regularized Matting Model

Based on the generated trimap, a novel L_0 regularized matting model is proposed to give an accurate segmentation of the OOI. It is observed that pixels which belong to the same object usually have similar color and are spatially close. Therefore, inspired by image matting algorithms, we used $Q(\boldsymbol{\alpha})$ to measure the smoothness of alpha matte $\boldsymbol{\alpha}$ with respect to the original image, and the similarity between $\boldsymbol{\alpha}$ and the trimap $\boldsymbol{\alpha}_s$.

$$Q(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T L \boldsymbol{\alpha} + \lambda(\boldsymbol{\alpha}^T - \boldsymbol{\alpha}_s^T) D_s (\boldsymbol{\alpha} - \boldsymbol{\alpha}_s), \quad (8)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_s$ are the vector form of alpha matte and the trimap, respectively, D_s is a diagonal matrix whose diagonal elements are one for labelled pixels in the trimap and zero for unlabelled pixels. λ is a parameter balancing the two terms, and is usually a large number to keep $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_s$ consistent at labelled locations. For a color image, L is a matting Laplacian matrix [10] whose (i,j)-th element is

$$\sum_{p|(i,j)\in T_p} \left(\delta_{ij} - \frac{1}{N_p} \left(1 + (I(i) - \mu_p)(C_p + \frac{\epsilon}{N_p} E_3)^{-1} (I(j) - \mu_p) \right) \right), \quad (9)$$

where δ_{ij} is the Kronecker delta function, $I(i)$ and $I(j)$ are the i-th and j-th pixel of original image I respectively. The summation is done for all p such that window T_p, which is the 3×3 window centered at p-th pixel, contains these two pixels. N_p is the number of pixels in this window, μ_p and C_p are the mean vector and covariance matrix of this window respectively. E_3 is the 3×3 identity matrix and ϵ is a regularizing parameter.

Moreover, in the case of segmentation, alpha values in $\boldsymbol{\alpha}$ should be either 1 or 0, which means the change of $\boldsymbol{\alpha}$ at the boundary of focused regions and defocused regions should be sharp and sparse. Besides, unsupervised OOI segmentation is sensitive to noise, because noise produces high frequency components which will greatly influence the identification of focused regions. To solve these problems, an L_0 norm prior $c(\boldsymbol{\alpha})$ is imposed during the estimation of $\boldsymbol{\alpha}$.

$$c(\boldsymbol{\alpha}) = \#\{p \big| \, |\partial_x \boldsymbol{\alpha}_p| + |\partial_y \boldsymbol{\alpha}_p| \neq 0\}, \quad (10)$$

where ∂_x and ∂_y denote the gradient along x and y directions. p is the pixel index and $\#\{\}$ is the counting operator which counts the number of p that satisfies $|\partial_x \boldsymbol{\alpha}_p| + |\partial_y \boldsymbol{\alpha}_p| \neq 0$. This is an L_0 gradient minimization problem. A similar problem is discussed in [11, 12] and it is shown that L_0 norm can globally control the number of non-zero gradients in order to constrain the sparse structure and reduce the impact of noise. Eqs. (8) and (10) are then fed into a unified optimization as

$$\min_{\boldsymbol{\alpha}} \ \big\{Q(\boldsymbol{\alpha}) + \gamma c(\boldsymbol{\alpha})\big\}, \tag{11}$$

where γ is a parameter that balances the two cost terms.

3. SOLVER

Since the L_0 optimization problem is highly non-convex, an axillary vector $\boldsymbol{\alpha}_0$ is introduced to relax the original problem stated in Eq. (11):

$$\min_{\boldsymbol{\alpha},\boldsymbol{\alpha}_0} \ \big\{Q(\boldsymbol{\alpha}) + \beta_1\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|_2^2 + \gamma c(\boldsymbol{\alpha}_0)\big\}. \tag{12}$$

When β_1 is large enough, $\boldsymbol{\alpha}_0$ perfectly approaches $\boldsymbol{\alpha}$. The original challenging problem is divided into two sub-problems, which are much easier to be solved. Satisfactory results can be obtained by iteratively solving the two sub-problems and increasing β_1 after each iteration. β_1 is initialized as a small value 0.01 and enlarged by a rate of 2 at each iteration. Vector $\boldsymbol{\alpha}_0$ is initialized as $\boldsymbol{\alpha}_s$.

3.1. Solve for $\boldsymbol{\alpha}$

Given $\boldsymbol{\alpha}_0$, the sparse prior term is constant and can be omitted in this part. Eq. (12) is rewritten as:

$$\min_{\boldsymbol{\alpha}} \ \big\{Q(\boldsymbol{\alpha}) + \beta_1\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|_2^2\big\}. \tag{13}$$

This is a quadratic optimization problem which has a closeform solution. $\boldsymbol{\alpha}$ is obtained by solving the following sparse linear system:

$$(L + \lambda D_s + \beta_1 E)\boldsymbol{\alpha} = \lambda D_s \boldsymbol{\alpha}_s + \beta_1 \boldsymbol{\alpha}_0, \tag{14}$$

where E is the identity matrix with the same size as L.

3.2. Solve for $\boldsymbol{\alpha}_0$

With $\boldsymbol{\alpha}$ fixed, the cost term $Q(\boldsymbol{\alpha})$ can be omitted and the second optimization problem is:

$$\min_{\boldsymbol{\alpha}_0} \ \big\{\beta_1\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|_2^2 + \gamma c(\boldsymbol{\alpha}_0)\big\}. \tag{15}$$

This is a non-convex problem, and two axillary vectors $\boldsymbol{h}$ and $\boldsymbol{v}$ are further employed to approximate $\partial_x\boldsymbol{\alpha}_0$ and $\partial_y\boldsymbol{\alpha}_0$, respectively.

$$\begin{aligned}\min_{\boldsymbol{\alpha}_0,\boldsymbol{h},\boldsymbol{v}} \ &\Big\{\beta_1\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|_2^2 + \gamma c(\boldsymbol{h},\boldsymbol{v}) \\ &+\beta_2(\|\partial_x\boldsymbol{\alpha}_0 - \boldsymbol{h}\|_2^2 + \|\partial_y\boldsymbol{\alpha}_0 - \boldsymbol{v}\|_2^2)\Big\},\end{aligned} \tag{16}$$

where $c(\boldsymbol{h},\boldsymbol{v}) = \#\big\{p\big|\ |\boldsymbol{h}_p| + |\boldsymbol{v}_p| \neq 0\big\}$, and p is the pixel index. Thus this problem is solved again with two sub-problems iteratively. Similar to Eq.(12), β_2 is increased after each iteration and satisfactory results can be obtained when β_2 is large enough. β_2 is initialized as $\beta_2 = 2\gamma$ and enlarged by a rate of 2 at each iteration. Vector $\boldsymbol{\alpha}_0$ is initialized as $\boldsymbol{\alpha}$ solved in Eq. (14).

3.2.1. *Solve for $\boldsymbol{h}$ and $\boldsymbol{v}$*

With $\boldsymbol{\alpha}_0$ fixed, Eq. (16) is simplified to

$$\min_{\boldsymbol{h},\boldsymbol{v}} \big\{\gamma c(\boldsymbol{h},\boldsymbol{v}) + \beta_2(\|\partial_x\boldsymbol{\alpha}_0 - \boldsymbol{h}\|_2^2 + \|\partial_y\boldsymbol{\alpha}_0 - \boldsymbol{v}\|_2^2)\big\}. \tag{17}$$

As this problem is independent of pixel index, a pixel-wise solution is obtained by

$$(\boldsymbol{h}_p,\boldsymbol{v}_p) = \begin{cases} (0,0), & \text{if } \|\partial_x\boldsymbol{\alpha}_{0p}\|_2^2 + \|\partial_y\boldsymbol{\alpha}_{0p}\|_2^2 \leq \gamma/\beta_2 \\ (\partial_x\boldsymbol{\alpha}_{0p}, \partial_y\boldsymbol{\alpha}_{0p}), & \text{otherwise} \end{cases}.$$

A large γ means the L_0 term is more important, and more non-zero gradient values will be set to zeros. The solution of Eq. (17) is similar to the work in [11] and a detailed proof can also be found.

3.2.2. *Solve for $\boldsymbol{\alpha}_0$*

Given $\boldsymbol{h}$ and $\boldsymbol{v}$, Eq. (16) is simplified to

$$\min_{\boldsymbol{\alpha_0}} \big\{\beta_1\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|_2^2 + \beta_2(\|\partial_x\boldsymbol{\alpha}_0 - \boldsymbol{h}\|_2^2 + \|\partial_y\boldsymbol{\alpha}_0 - \boldsymbol{v}\|_2^2)\big\}. \tag{18}$$

This is a quadratic optimization problem, and the solution is easily obtained after Fourier Transform.

$$\boldsymbol{\alpha}_0 = \mathcal{F}^{-1}\left(\frac{\beta_1\mathcal{F}(\alpha) + \beta_2(\mathcal{F}(\partial_x)^*\mathcal{F}(h) + \mathcal{F}(\partial_y)^*\mathcal{F}(v))}{\beta_1\mathcal{F}(1) + \beta_2(\mathcal{F}(\partial_x)^*\mathcal{F}(\partial_x) + \mathcal{F}(\partial_y)^*\mathcal{F}(\partial_y))}\right),$$

where $\mathcal{F}$ is Fourier Transform operator and $*$ denotes complex conjugate. $\mathcal{F}(1)$ is the Fourier Transform of delta function. The plus, multiplication and division are all pixel-wise operators.

4. EXPERIMENTAL RESULTS

4.1. Experiment Setup and Parameter Setting

Experiments were conducted on a dataset build from BSDS500 [13]. 86 images with obvious low DOF are chosen from BSDS500 to evaluate the precision, robustness and efficiency of the proposed method. The proposed method is compared with two unsupervised low DOF segmentation methods which generate the state-of-the-art results [6, 8]. Codes of these two methods are obtained online, and the default parameters provided by the authors are adopted to keep consistency with the results given in the original papers. Since the ground truth segmentation results are available in BSDS500,

(a) Original image (b) Ground truth (c) [6] (F=0.9295) (d) [8] (F=0.9322) (e) Proposed (F=0.9838)

Fig. 1. Segmentation results of 189011.

(a) Original image (b) Ground truth (c) [6] (F=0.7724) (d) [8] (F=0.8074) (e) Proposed (F=0.9213)

Fig. 2. Segmentation results of 42049.

F measure is chosen to evaluate the performance of different methods. In general, larger F value means bigger portion of correct result, indicating a better segmentation result.

The parameters used in the proposed method are set as follows. The parameters of guided filter in Eq.(4) are set to be $r = 7$ and $\xi = 10^{-6}$. In Eq.(6), a disk with radius $r_d = 5$ is used as the kernel of dilation, and a disk with radius $r_e = 7$ is used as the kernel of erosion. λ in Eq.(8) is 10 and γ in Eq.(11) is 0.04.

Table 1. Performance Comparison of Different Methods

Index	[6]	[8]	Proposed Method
max(F)	0.9863	0.9875	0.9893
min(F)	0.3541	0.3771	0.5130
average(F)	0.7709	0.8048	0.8369
std(F)	0.1658	0.1406	0.1163
number of best results	5	13	68
average computation time per image	110s	420s	42.8s

4.2. Results Comparison

Fig.1 illustrates the segmentation results of one test image (189011) in our dataset. The original image and ground truth segmentation results are shown in Fig.1a and Fig.1b, respectively. The results of Graf's method [6] and Mei's method [8] are shown in Fig.1c and Fig.1d, respectively. Fig.1e is the result of the proposed method. It can be seen that the proposed method finds the exact boundary of focused regions and defocused regions, of which the outline of straw hat is a good example. Moreover, the proposed method has good performance in smooth regions. Existing methods have difficulties in segmenting the basket in the boy's hands since little texture appears on the surface of basket. The proposed method perfectly segments the basket because the matting terms in Eq.(8) provide a good spreading effect. Besides, the proposed method is robust to noise. In Fig.1c and Fig.1d, some isolated spots in the defocused regions are wrongly segmented due to the existence of noise, however no such spots appear in Fig.1e. Another example is demonstrated in Fig.2, where the good performance of the proposed method and robustness to noise are also demonstrated.

The performance of the proposed method and the other methods over the whole test dataset are presented in Table 1. It is shown that the proposed method generally outperforms the other methods. Besides, the proposed method requires less computation time to generate satisfactory results.

5. CONCLUSION

We have proposed an effective unsupervised segmentation method to extract the object of interest from low depth of field images based on a novel L_0 regularized matting model. The guided filter and morphological filter after the multi-scale reblurring process effectively eliminate the scattered error points and generate a satisfactory trimap. The introduction of matting terms leads to good performance in smooth regions and helps find the accurate boundary of focused regions and defocused regions. The adoption of L_0 norm reduces artifacts in boundary regions and improves robustness to noise.

6. REFERENCES

[1] Thomas Meier and King N Ngan, "Video segmentation for content-based coding," *IEEE transactions on circuits and systems for video technology*, vol. 9, no. 8, pp. 1190–1203, 1999.

[2] S Kavitha, S Mohammed Mansoor Roomi, and N Ramaraj, "Lossy compression through segmentation on low depth-of-field images," *Digital Signal Processing*, vol. 19, no. 1, pp. 59–65, 2009.

[3] Akira Kubota and Kiyoharu Aizawa, "Reconstructing arbitrarily focused images from two differently focused images using linear filters," *ieee Transactions on Image Processing*, vol. 14, no. 11, pp. 1848–1859, 2005.

[4] Du-Ming Tsai and Hu-Jong Wang, "Segmenting focused objects in complex visual images," *Pattern Recognition Letters*, vol. 19, no. 10, pp. 929–940, 1998.

[5] James Ze Wang, Jia Li, Robert M Gray, and Gio Wiederhold, "Unsupervised multiresolution segmentation for images with low depth of field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 1, pp. 85–90, 2001.

[6] Franz Graf, Hans-Peter Kriegel, and Michael Weiler, "Robust segmentation of relevant regions in low depth of field images," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 2861–2864.

[7] Hongliang Li and King N Ngan, "Unsupervized video segmentation with low depth of field," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1742–1751, 2007.

[8] Jiangyuan Mei, Yulin Si, and Huijun Gao, "A curve evolution approach for unsupervised segmentation of images with low depth of field," *IEEE Transactions on image processing*, vol. 22, no. 10, pp. 4086–4095, 2013.

[9] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *European conference on computer vision*. Springer, 2010, pp. 1–14.

[10] Anat Levin, Dani Lischinski, and Yair Weiss, "A closed-form solution to natural image matting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228–242, 2008.

[11] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia, "Image smoothing via l 0 gradient minimization," in *ACM Transactions on Graphics (TOG)*. ACM, 2011, vol. 30, p. 174.

[12] Shuai Yi, Xiaogang Wang, Cewu Lu, and Jiaya Jia, "L0 regularized stationary time estimation for crowd group analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2211–2218.

[13] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 416–423.