Robust Multi-Focus Image Fusion Using Edge Model and Multi-Matting

Yibo Chen, Member, IEEE, Jingwei Guan, Member, IEEE, and Wai-Kuen Cham, Senior Member, IEEE

Abstract—An effective multi-focus image fusion method is proposed to generate an all-in-focus image with all objects in focus by merging multiple images. The proposed method first estimates focus maps using a novel combination of edge model and a traditional block-based focus measure. Then, a propagation process is conducted to obtain accurate weight maps based on a novel multi-matting model that makes full use of the spatial information. The fused all-in-focus image is finally generated based on a weighted-sum strategy. Experimental results demonstrate that the proposed method has state-of-the-art performance for multi-focus image fusion under various situations encountered in practice, even in cases with obvious misregistration.

Index Terms—Multi-focus, fusion, edge model, multi-matting.

I. INTRODUCTION

MULTI-FOCUS image fusion is an important technique to tackle the defocus problem that some parts of the image are out of focus and look blurry. This problem usually happens when a low depth-of-field optical system is used to capture an image, such as microscopes or large aperture cameras. Through multi-focus image fusion, several images of the same scene but with different focus settings can be combined into a so-called all-in-focus image, in which all parts are fully focused. A good multi-focus image fusion method should meet the requirements that all information of the focused regions in each source image is preserved in the fused image with little artifacts, and the method is robust to imperfect situations such as existence of noise and misregistration.

During the last decade, a large number of multi-focus image fusion methods have been proposed. Most of them can be categorized into four major groups [1]. The first group consists of methods based on multi-scale decomposition [2], such as pyramid [3], wavelet decomposition [4], [5], curvelet transform [6], shearlet transform [7], non-subsampled contourlet transform (NSCT) [8], [9] and neighbor distance (ND) [10]. Recently guided filter was also introduced to refine the multi-scale representations and the method achieves state-of-the-art performance [11]. The second group consists of sparse representation (SR) based methods. The concept of SR was first introduced into image fusion in [12], which employs the orthogonal matching pursuit algorithm to image patches of multiple source images. Adaptive SR model was proposed in [13] for simultaneous image fusion and denoising. In [14], an over-complete dictionary was learned from numerous training samples similar to the source images, thus adding adaptability for the sparse representation. Zhang et al. [15] also proposed a multi-task SR method that considers not only sharpness information, like edges and textures, of each image patch but also relationship among source images. The major advantage of these two groups of methods is that sharpness information can be well preserved in the fused image. However, since spatial consistency is not fully considered, brightness and color distortions may be observed. Moreover, slight image misregistration will result in ghosting artifacts due to their sensitivity to high-frequency components.

The third group consists of methods based on computational photography techniques such as light field rendering [16]. This kind of methods discovers more of the physical formation of multi-focus images and reconstructs the all-in-focus images.

The last group consists of methods performed in spatial domain, which can make full use of the spatial context and provide spatial consistency. A commonly used approach is to take each pixel in the fused image as the weighted average of the corresponding pixels in the source images. The weights are determined according to how well each pixel is focused, which is usually measured by some metrics. These metrics, such as spatial frequency inside a surrounding block, are named as focus measures. However, this kind of methods may generate blocking artifacts on object boundaries due to the block-based computation of focus measure. To solve this problem, some region-based fusion methods have been proposed and calculation is done in regions with irregular shapes obtained by segmentation [17]. A common limitation of region-based methods is that they rely heavily on the accuracy of segmentation. Another way is post-processing of the initial weight maps [18]–[20]. In [21], a noise-robust selective fusion method that models the change of weights as a Gaussian function is proposed, but it can only achieve good performance on properly ordered image sequence with 3 or more images. Recently, image matting was introduced to refine the initial weight maps and can achieve good performance [22]. However, the performance relies on good initial weight maps, and the common block-wise computation is still a limitation. Our previous work [23] uses one of the parameters of edge model directly as focus measure. However, edge model is a
sparse feature representation method and focus measure values only exist at edge locations. So image matting was applied to propagate the focus measure values to the whole image. Since the matting process handles source images separately without considering the correlation among them, the propagation effect is very limited and the fusion process is also not stable in many situations.

To solve the problems mentioned above, a novel edge model and multi-matting (EMAM) based multi-focus image fusion method is proposed in this paper, which can be classified as the fourth group of methods. Firstly, a novel method that combines a traditional focus measure and edge model is developed to estimate focus maps. Then a new multi-matting model is proposed to spread the focus maps to the whole image region considering the correlation among source images and obtain final weight maps. The proposed method maintains the advantages of the last group of methods and meanwhile overcomes the weakness. Experimental results demonstrate that the proposed method outperforms current state-of-the-art fusion methods under various situations. The main contributions of the proposed method outperforms current state-of-the-art fusion methods. Firstly, a novel method that combines the matting process handles source images separately without only exist at edge locations. So image matting was applied to sparse feature representation method and focus measure values.

B. Edge Modeling

To exploit the information of above defocus model from a defocused image, a parametric edge model [26]–[29] is adopted here. As shown in Fig. 1(a), edges in a 2D image have local 1D structure features, where sharp intensity changes can be observed in the direction of $x$ and nearly no change in the perpendicular direction. Consider an edge point which locates at $x_0$ as shown in Fig. 1(a). The intensity in the direction of $x$, as represented by $s_1(x)$ in Fig. 1(b), is modeled as a Gaussian smoothed step edge which is equal to the convolution of a step edge and a Gaussian filter:

$$s_1(x; b, c, x_0) = e(x; b, c, x_0) \otimes h(x; \sigma_1),$$

where $e(x; b, c, x_0) = cU(x - x_0) + b$ and $h(x; \sigma_1) = (1/\sqrt{2\pi\sigma_1^2})\exp(-x^2/2\sigma_1^2)$. $U(\cdot)$ is unit step function, $b$ represents edge contrast, $c$ represents edge basis and the spread parameter of Gaussian filter $\sigma_1$ represents edge width.

An existing fitting process is utilized to extract edge parameters $\sigma_1, x_0, b$ and $c$. The typical edge $s_1(x; b, c, x_0)$ is convolved with the derivative of a Gaussian filter $h'_d(x; \sigma_d)$, and the output shown in Fig. 1(c) is denoted as $d(x; c, \sigma_1, x_0, \sigma_d)$. Then the edge parameters $\sigma_1$ and $x_0$ are estimated:

$$\sigma_1 = \sqrt{\frac{1}{\ln(d_1/d_2^d)} - \sigma_d^2}, \quad x_0 = \frac{\ln(d_1)}{2\ln(d_1/d_2^d)},$$

where $d_1 = d(x_c; c, \sigma_1, x_0, \sigma_d)$, $d_2 = d(x_c + 1; c, \sigma_1, x_0, \sigma_d)$ and $d_3 = d(x_c - 1; c, \sigma_1, x_0, \sigma_d)$. $x_c$ is the discrete location with largest $d(x; c, \sigma_1, x_0, \sigma_d)$ value. Empirical results suggest that $\sigma_d = 1$ is a reasonable value for most natural images.

According to (1), an edge in a defocused region can be regarded as the convolution of a typical edge and a Gaussian filter with spread parameter $\sigma_f$. Since the convolution of two Gaussian functions is still Gaussian, the defocused edge $s_2(x)$ at location $x_0$, as shown in Fig. 1(d), can be represented as:

$$s_2(x; x_0) = s_1(x; b, c, x_0) \otimes h(x; \sigma_f)$$

$$= e(x; b, c, x_0) \otimes h(x; \sigma_1) \otimes h(x; \sigma_f)$$

$$= e(x; b, c, x_0) \otimes h(x; \sigma_2),$$
where $\sigma_2 = \sqrt{\sigma_1^2 + \sigma_f^2}$. The defocused edge $s_2(x)$ also fits the edge model and $\sigma_2$ can be estimated by (4). Obviously $\sigma_2$ is larger than $\sigma_1$.

To determine the direction of the edge at each pixel, gradients along the two image axis directions are first estimated separately. Then the perpendicular direction of the edge can be obtained by the ratio of these two gradients and edge model is applied to estimate the parameters along this direction.

### III. Problem Formulation and Proposed Approach

Multi-focus image fusion is to fuse several source images, which are captured with different focus settings, into an all-in-focus image. The source images are usually assumed to be well aligned. In general, most multi-focus fusion methods performed in spatial domain contain two steps: focused region identification and focus map estimation, weight map estimation and fusion. Each source image is only well focused in some local regions. To identify these focused regions, focus measure is usually used to make comparisons among source images at the same location and to estimate a focus map for each source image. Then a fused image is obtained as a weighted sum of the source images, where the needed weight maps can be estimated from the focus maps.

Our proposed method follows the general steps of methods performed in spatial domain. Fig. 2 summarizes the main processes of the proposed fusion method. First, initial focus maps are estimated using a traditional focus measure, and the edge model described in section II-B is utilized to refine them. Then a multi-matting scheme is used to obtain accurate focused regions from the focus maps and to estimate weight maps for all source images. With these weight maps, the fused all-in-focus image $I_F$ is obtained as a weighted sum of the source images $I_1, \ldots, I_N$:

$$I_F(i) = \sum_{k=1}^{N} w_k(i) I_k(i),$$

where $i$ denotes pixel index.

### A. Focus Map Estimation and Refinement

Regions labelled in a focus map represent the best focused regions among all source images. Since regions wrongly labelled as the best focused regions of an source image will probably result in artifacts in the final fused image, it is much less damaging if no source image is identified as the best focused one in these unreliable regions. The proposed method achieves good fusion results even though only parts of focused regions are labelled in each focus map by means of multi-matting. In this paper, a novel method which combines a traditional focus measure and edge model is proposed to estimate reliable focus maps.

A comparison of various traditional focus measures can be found in [30] and [31]. As shown in these papers, gradient energy [32] is a good candidate that provides high precision. The focus measure of a pixel at location $(x, y)$ in the $k$-th source image $I_k(k = 1, \cdots, N)$ is defined as:

$$FM_k(x, y) = \sum_{(\hat{x}, \hat{y}) \in \Omega(x, y)} (I_{kx}(\hat{x}, \hat{y})^2 + I_{ky}(\hat{x}, \hat{y})^2),$$

where $\Omega(x, y)$ is the $r \times r$ block with $(x, y)$ at the center, $I_{kx}$ and $I_{ky}$ denote the gradient values in vertical and horizontal directions in source image $I_k$ respectively. The source image that is best focused at $(x, y)$ should have the largest focus measure value among all source images at the same location.

At each location $(x, y)$, focus measures $FM_k(x, y)(k = 1, \cdots, N)$ corresponding to the $N$ source images are compared with each other, and the largest focus measure $FM_t(x, y)$ is found. Then two conditions are used to verify that the decision of $t$ is reliable, which implies source image $I_t$ is definitely best focused at $(x, y)$.

$$FM_t(x, y) > FM_{th},$$

$$\frac{1}{N-1} \sum_{k \neq t, k=1, \cdots, N} FM_k(x, y) > R_{th},$$

where $FM_{th}$ is a constant threshold for the maximum focus measure value, and $R_{th}$ is a constant threshold for the ratio of the maximum focus measure to the mean of other focus measures at each location. Then initial focus maps $M_k(k = 1, \cdots, N)$ can be generated as follows.

$$M_k(x, y) = \begin{cases} 1 & \text{if } k = t, \\ 0 & \text{if } k \neq t \text{ or } t \text{ is not reliable}. \end{cases}$$
It should be noted that at some locations no source image satisfies conditions (8) and (9). Thus, the decision of t is regarded as not reliable and all focus maps will be set to zeros at these locations. One reason for these two conditions is based on observations of wrong decisions in flat regions which have little texture. For each location in these regions, the focus measures of source images generally have low magnitude and low inter-image contrast. The decision of t will be unstable and therefore (8) and (9) are introduced.

Another reason is the limitation of block computation, which is demonstrated in Fig. 3. Source image 1 in Fig. 3a focuses in background (a big clock) and source image 2 in Fig. 3b focuses in foreground (a small clock). Since focus measure of each pixel is computed from all gradients inside an \( r \times r \) block centered at the pixel, problems may arise at the boundary of foreground and background. This is especially true when the block centered at a pixel contains both rich foreground texture and rich background texture, such as pixel B and pixel \( B' \) in Fig. 3. B and \( B' \) are actually foreground pixels and the focus measure of \( B' \) should be larger than that of B since \( B' \) is better focused. But the sharp edges of shape ‘8’ in the background, which are inside the block centered at B, greatly influence the computation of focus measure. Thus, the focus measure of \( B' \) may be very similar to B, which will make the comparison result unreliable. This problem does not happen for pixels A and \( A' \), which are also at a same location, since the background inside the block centered at that location is flat with little texture and will not influence focus measure computation. Fig. 4 shows the initial focus maps of the two source images from Fig. 3, where only reliable pixels such as A are marked and confusing pixels such as B will be ignored by conditions (8) and (9).

However, these confusing pixels carry important information which can be used as a guidance at the boundary of foreground and background, and it is unwise to simply ignore them. That is why edge model is introduced. Since edge width \( \sigma \) and edge location \( x_0 \) are calculated using only a few nearby pixels instead of a block, as stated in Section II-B, they represent the edges at the boundary more precisely and complement the ignored information. However, estimation of edge model parameters may not be accurate in regions where edges are too serried and irregular, in which case the gradient energy focus measure performs better. Besides, direct comparison of edge width \( \sigma \) is sensitive to the location shift of edges among source images, especially when the source images are not perfectly aligned. Therefore edge model and the gradient energy focus measure are combined to complement each other by following steps.

Firstly, edge width \( \sigma \) and edge location \( x_0 \) are calculated for each pixel based on (4). Then an edge map for each source image is obtained with values of edge width at corresponding edge locations. However, estimation of edge location \( x_0 \) will be inaccurate when edge is heavily defocused. In other words, edge locations calculated from edges with large edge width values are not reliable. It is shown in [26] that for most well-captured natural images, clear edges have edge width \( \sigma \) ranging from 0.01 to 0.5. Thus for each edge map, only edge locations with edge width smaller than \( \sigma_{th} = 0.5 \) are remained.

Secondly, an edge linking algorithm [33] is applied to refine the edge maps and all connected edge contours in each edge map are listed. Finally, for each edge contour, we identify the source image in which this contour is more likely to be focused. Then refined focus maps are obtained, which are shown in Fig. 5. If the focus maps have some intersections, though only at very few locations, we will eliminate these locations from all focus maps to ensure the focus maps reliable. Edge model gives an instruction in those confusing regions, refines the initial focus maps, and greatly benefits the following multi-matting process.

Edge model is used instead of Canny edge detector because edge model provides parameter \( \sigma \) in addition to edge contour. Information of \( \sigma \) plays an important role in the proposed algorithm and cannot be replaced by the threshold of Canny edge detector. As described in Section II, defocus process is modeled as Gaussian blur. Edge model is derived from this defocus model and parameter \( \sigma \) perfectly represents the degree of defocus, uninfluenced by edge contrast. However, Canny edge detector uses gradient as measure, which is easily
influenced by edge contrast, and setting a threshold for Canny edge’s measure is not accurate when examining the degree of defocus. Thus, the edge map produced by edge model is more reliable and fit for our purpose.

B. Weight Map Estimation With Multi-Matting
Model and Fusion

Since only parts of focused regions are labelled in each focus map after the above focus map estimation, a propagation step is needed to spread them based on local smoothness assumptions of colors. Then weight maps are obtained and used to generate the fused all-in-focus image. Image matting algorithm is a good choice which is adapted to meet our demand.

1) Image Matting and Application in Multi-Focus Fusion:

Image matting aims at extracting foreground and background from an observed image $I_o$. Foreground and background will be represented as layers $I_{p1}$ and $I_{p2}$ respectively. The color of $i$-th pixel in $I_o$ is assumed to be a linear combination of the colors in the two layers

$$I_o(i) = \alpha(i) I_{p1}(i) + (1 - \alpha(i)) I_{p2}(i), \quad (11)$$

where $\alpha$ is the opacity of layer $I_{p1}$ called alpha matte. $\alpha(i) = 1$ or 0 means the $i$-th pixel belongs to layer $I_{p1}$ or layer $I_{p2}$, respectively. Image matting is to obtain an accurate alpha matte $\alpha$ together with the two layers. Since solving the three unknowns $\alpha$, $I_{p1}$, $I_{p2}$ in (11) from a single image $I_o$ is an underconstrained problem, additional information is required to extract a good matte. Most recent methods [34]–[36] use a trimap which roughly segments the image into three regions: layer $I_{p1}$, layer $I_{p2}$ and unknown.

Among all image matting algorithms, the closed-form matting method by Levin [36] is a good choice because of its high accuracy and low requirement of trimap. This method is based on local smoothness assumptions on colors of the two layers, which also stand for the case of focus map. The alpha matte is obtained by solving

$$\arg\min_\alpha \alpha^T L \alpha + \lambda (\alpha^T - \alpha_o^T) D_s (\alpha - \alpha_s), \quad (12)$$

where $\alpha_s$ is the vector form of the trimap and $D_s$ is a diagonal matrix whose diagonal elements are one for labelled pixels which have specified alpha values in the trimap and zero for all the other pixels. $\lambda$ is a parameter that controls the similarity between alpha matte $\alpha$ and the trimap $\alpha_s$ at labelled locations. Usually $\lambda$ is a large number to keep them consistent at labelled locations. For a color image, $L$ is a matting Laplacian matrix whose $(i,j)$th element is

$$\sum_{p((i,j)\in T_p)} \left( \delta_{ij} - \frac{1}{N_p} \left( 1 + (I(i) - \mu_p)(C_p + \epsilon N_p E_3)^{-1} (I(j) - \mu_p) \right) \right), \quad (13)$$

where $\delta_{ij}$ is the Kronecker delta function, $I(i)$ is the $i$-th pixel of image $I$ and $I(j)$ is the $j$-th pixel of image $I$. The summation is done for all windows $T_p$ that contain these two pixels. $T_p$ is the $3 \times 3$ window centered at $p$-th pixel, $N_p$ is the number of pixels in this window, $\mu_p$ and $C_p$ are the mean vector and covariance matrix of this window respectively. $E_3$ is the $3 \times 3$ identity matrix and $\epsilon$ is a regularizing parameter. Since the cost function (12) is quadratic, $\alpha$ is obtained by solving the following sparse linear system:

$$(L + \lambda D_s) \alpha = \lambda D_s \alpha_s. \quad (14)$$

Image matting algorithms can be used to obtain accurate focused region in image and video applications [37], [38]. More specifically, for each source image, regions which are best focused among all source images are regarded as layer $I_{p1}$, and the remaining regions are all regarded as layer $I_{p2}$. The refined focus map obtained in Section III-A can be used as the trimap. The resulting alpha matte obtained by the matting algorithm above can discriminate needed best focused regions from each source image, and the alpha matte is regarded as an initial weight map.

However, this image matting method can only handle a single image at one time. In the case of multi-focus fusion, two or more focus maps need to be handled, each of which corresponds to one of the source images. A common idea is to separately handle each focus map using the abovementioned single image matting and then normalize the generated initial weight maps to make them sum up to one at each location. Then these normalized weight maps are used to obtain multi-focus fusion result by the weighted sum of source images. However, this idea has problems when all initial weight maps have very small values at some locations. Errors will be magnified at these locations and the fusion result will not be stable. Besides, source images are actually capturing the same scene and must have high correlation, which is also not fully utilized. Another similar method can be found in [22], which uses single image matting to obtain initial weight maps and generates final weight maps using a non-linear combination of these initial weight maps. However, source images are not treated with equal importance under the non-linear operation, and different order of source images will lead to different fusion result. Besides, one of the focus maps is not used in the non-linear operation, which will cause information lost and errors possibly. Moreover, it cannot solve the problem when all initial weight maps have very small values at some locations either.

2) Proposed Multi-Matting Model: In consideration of the limitations of existing matting-based fusion methods and the high correlation among source images, a novel multi-matting model is proposed to estimate weight maps from the focus maps.

The focused region of one source image $I_k$ will spread out when this region is defocused in other source images. So the focused region can only be accurately extracted from $I_k$. Then a cost term of each weight map is defined using only the local smoothness constraints derived from $I_k$.

$$E_s(w_k) = w_k^T L_k w_k + \lambda_k (w_k^T - m_k^T) D_k (w_k - m_k), \quad (15)$$

where $m_k$ is the vector form of the $k$-th focus map $M_k$, and $D_k$ is a diagonal matrix whose diagonal elements are one for labelled pixels in $m_k$ and zero for all other pixels. $w_k$ is the vector form of the $k$-th weight map to be solved, and $L_k$ is the
matting Laplacian matrix of the $k$-th source image $I_k$ defined in the same way as (13). $\lambda_k$ is a parameter that controls the similarity between weight map $w_k$ and the focus map $m_k$ at labelled locations.

To utilize the correlation among source images, the cost terms of all weight maps are taken into consideration in a single optimization problem under a strong constraint that the sum of weight maps at each location should be equal to one. The parameters $\lambda_1, \ldots, \lambda_N$ are all set to the same $\lambda$ to ensure the equal importance of all focus maps.

$$\arg\min_{w_1, \ldots, w_N} \sum_{k=1}^{N} \left( w_k^T L_k w_k + \lambda (w_k^T - m_k^T) D_k (w_k - m_k) \right),$$

subject to $\sum_{k=1}^{N} w_k = e$, \hspace{1cm} (16)

where $e$ is an all-ones vector. It should be noted that existing matting-based methods need to generate a trimap for each focus map with both foreground labels and background labels. However, the proposed method only needs foreground labels on each focus map $m_k$ and gives more freedom to the propagation of focus maps.

Solving (16) is a quadratic programming problem with only equality constraints. The Lagrange function and lagrange multipliers are used, and it is readily shown that the solution is given by the following linear system:

$$\begin{bmatrix} L_1 + \lambda D_1 & 0 & 0 & E \\ 0 & \ddots & 0 & E \\ 0 & 0 & L_N + \lambda D_N & E \\ E & E & E & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_N \\ \theta/2 \end{bmatrix} = \begin{bmatrix} \lambda m_1 \\ \vdots \\ \lambda m_N \\ e \end{bmatrix},$$

(17)

where $E$ is the identity matrix with the same size as $L_k$ and $\theta$ is a set of Lagrange multipliers. Since the matting Laplacian matrixes $L_k$ defined by (13) are symmetric and very sparse, the leftmost matrix in (17) is also symmetric and sparse. Thus the close-form solution to this optimization problem can be computed with not too much computation cost.

3) Comparison With Single Image Matting Method:

In single image matting, the matte $w_0k$ of image $I_k$ is obtained by:

$$w_{0k} = (L_k + \lambda D_k)^{-1} \lambda D_k m_k.$$  

If we denote the sum of all the mattes of source images as

$$w_{0A} = \sum_{k=1}^{N} ((L_k + \lambda D_k)^{-1} \lambda D_k m_k),$$

then the matte for each source image after normalization is:

$$\bar{w}_{0k} = w_{0k} / w_{0A} = w_{0k} + (-w_{0k} / w_{0A}) \ast (w_{0A} - e),$$

(18)

where $/ \ast$ and $\ast$ denote pixel-wise division and multiplication, respectively.

In multi-matting, the solution of (17) can be obtained through matrix computation, which is detailed in Appendix. The matte for each source image after multi-matting can be expressed as:

$$w_k = w_{0k} + (L_k + \lambda D_k)^{-1} S_{ch}^{-1} \left( w_{0A} - e \right),$$

(19)

where $S_{ch}$ is the Schur complement of the left coefficient matrix in (17)

$$S_{ch} = - \sum_{k=1}^{N} (L_k + \lambda D_k)^{-1}.$$  

From the comparison of (18) and (19), it can be seen that the single image matting method only uses each matte $w_{0k}$ and $w_{0A}$ for normalization, and this process is not stable since $w_{0A}$ may be very close to 0 at some locations. The proposed multi-matting method makes more use of the correlation among source images implied by $L_k, D_k$ and $S_{ch}$, thus provides better guidance in regions where single image matting performs unsatisfactory.

The constraint in the optimization problem is derived from, but is not limited to weight normalization. It not only provides an adaptive normalization according to image contents, but also provides the correlation among source images to the matting based cost function. Existing matting based fusion methods conduct focus map matting and weight map normalization separately, and ignore the fact that these two processes can help each other. However, the proposed method utilizes the fact and achieves better performance, especially in flat regions and boundary regions.

In flat regions, where the focus maps have few labels, it is hard to have a good spread effect here using single image matting. The generated initial weight maps will all have small values at these locations and errors will be magnified during normalization. However, the proposed method realizes an adaptive normalization by minimizing the matting cost function. It helps judging which source image is more likely to be best focused at these locations based on smoothness of image content.

In boundary regions, where the initial focus maps have many labels, the single image matting may spread too much to unwanted regions. However, the equality constraint in the proposed method explores the correlation among source images. That is, at each location, if the weight value of one weight map is close to 1, the other weight maps should have very small weight values close to 0. Thus the excessive spreading will be suppressed when minimizing the matting based cost function.

4) Iterative Refinement: Recall that weight maps should be close to 0 or 1 for most image pixels since majority of image pixels are opaque and can be found best focused in one of the source images. This kind of constraints can be modeled using $L_1$ norm or integer programming in optimization, both of which are hard to be solved [39]. However, by observing results of image matting methods, we found that the pixels which are far from the labelled pixels are not likely to have alpha values close to 0 or 1 even though they have similar color with the labelled pixels. In other words, quality of the initial focus maps is important, and more labelled pixels will lead to better results.
Algorithm 1 Multi-Matting Model for Multi-Focus Fusion

1: Input: parameter $\lambda$, focus maps $m_k$, label indication matrices $D_k$, matting Laplacian matrices $L_k$, $k = 1, \cdots, N$.
2: Output: weight maps $w_k$, $k = 1, \cdots, N$.
3: Initialize: Solve for $w^{(0)}$ in (17) and $E_b^{(0)}$ in (21), $\varepsilon = 0.01$.
4: while not converged do
5:   With $w^{(q)}$, compute the two thresholds $t_{k1}^{(q)}$, $t_{k2}^{(q)}$ using multi Otsu method.
6:   With $w^{(q)}$, $t_{k1}^{(q)}$, and $t_{k2}^{(q)}$, solve for $m_k^{(q)}$ in (20).
7:   With $m_k^{(q)}$, obtain $D_k^{(q)}$.
8:   With $m_k^{(q)}$ and $D_k^{(q)}$, solve for $w^{(q+1)}$ in (17) and $E^{(q+1)}$ in (21).
9:   Check the convergence conditions:
   \[ |E^{(q+1)} - E^{(q)}| / E^{(q)} < \varepsilon \]
10: $q = q + 1$.
11: end while

An iterative scheme is used to update $m_k$ by adding more labelled pixels, and then the updated $m_k$ maps are taken back to (17) to solve for more accurate weight maps. New labelled pixels are extracted from $w_k$ by multilevel thresholding algorithm, such as the multi Otsu method [40]. This classification algorithm calculates the optimum thresholds separating the pixels into several classes so that their intra-class variance is minimal. Thus, by setting up two thresholds, pixels of $w_k$ are divided into three classes, near-one pixels, near-zero pixels and intermediate uncertain pixels. Then all pixels in the near-one class are set to be 1, all pixels in the near-zero class are set to be zeros, while the other pixels remain unlabelled. Thus a new set of $m_k$ is generated and the iteration goes on.

\[
m_k^{(q)}(i) = \begin{cases} 
1 & \text{if } w_k^{(q)}(i) > t_{k1}^{(q)} \\
0 & \text{if } w_k^{(q)}(i) < t_{k2}^{(q)} \\
\text{unlabelled otherwise}, & \text{if } t_{k1}^{(q)} < t_{k2}^{(q)}.
\end{cases} \tag{20}
\]

where $t_{k1}^{(q)}$ and $t_{k2}^{(q)}$ ($t_{k1}^{(q)} < t_{k2}^{(q)}$) are the two thresholds obtained from multi Otsu algorithm in the $q$-th iteration, and $i$ denotes the index of pixel. Also, a decision term is defined to decide when to terminate the iteration process.

\[
E_b = \sum_k \|w_k(1 - w_k)\|_1 = \sum_k \|w_k(i)(1 - w_k(i))\|_1. \tag{21}
\]

where $\|\cdot\|_1$ denotes $L_1$ norm of a vector, and the absolute value $|w_k(i)(1 - w_k(i))|$ will be smaller if $w_k(i)$ is closer to 0 or 1. The value of $E_b$ becomes smaller after each iteration, and the iteration will be stopped when $E_b$ doesn’t change much. The proposed algorithm is summarized in Algorithm 1. Usually results after 3 to 4 iterations are already good enough for the case of multi-focus fusion.

The weight maps using another matting based method IM in [22] and the proposed method are shown in Fig. 6b and 6d, respectively. To further demonstrate the performance of multi-matting, we replaced the multi-matting part with single image matting and normalization in each iteration of the proposed method, and the result is shown in Fig. 6c. Two regions are selected to illustrate the benefit of multi-matting. The red box is a boundary region of foreground and background. It is obvious that the proposed method suppresses the excessive spreading and gives a more exact location of the boundary. The blue circle is a background region with a flat region on the left and a boundary of the ‘clock’ on the right. It is hard for single image matting to extract the flat region with good accuracy and the errors cannot be eliminated in the iterative refinement. The weight map values relate to the proportion of the corresponding source image in the final fused image. Since the background of the source image in Fig. 6a is heavily blurred, the weight maps shown in Fig. 6b and 6c will produce blur at the boundary of the ‘clock’ in background. In summary, multi-matting generates exact location of the boundary and iterative refinement makes the weight maps close to 0 or 1.

It should be noted that for color source images, we use luminance values to estimate traditional focus measure and edge model parameters as described in Section III-A. Then RGB values are used in multi-matting model as described in Section III-B because RGB values can provide more information for image matting than luminance values.

IV. EXPERIMENT RESULTS AND DISCUSSIONS

A. Experimental Setup

Experiments were conducted on 36 sets of multi-focus images, all of which are publicly available. 20 sets of them come from a new multi-focus dataset called “Lytro” which is publicly available online [41]. The other 16 sets are collected...
online, which consist of one set of medical multi-focus images “bug” used in “Digital Photomontage” project [42], one set of Infrared multi-focus images “IR1g”, and another 14 sets which are widely used in multi-focus fusion research. The diversity of the image sets provides a good representation of various situations encountered in practice. A portion of the image sets are shown in Fig. 7.

The proposed EMAM fusion method is compared with six multi-focus image fusion algorithms which are non-subsampled contourlet transform (NSCT) [8], guided filtering (GF) [11], image matting (IM) [22], sparse representation (SR) [12], non-subsampled contourlet transform and sparse representation (NSCT-SR) [43], and dense scale invariant feature transform (DSIFT) [20]. Codes of all these methods are obtained online or directly from the authors, and the default parameters provided by the authors are adopted to keep consistency with the results given in the original papers.

B. Objective Image Fusion Quality Metrics

To evaluate the performance of different fusion methods objectively, several fusion quality metrics are needed. Since ground truth results for the case of multi-focus image fusion are usually not available, non-referenced quality metrics are generally preferred. A good survey of these kind of metrics can be found in [44]. Four quality metrics are adopted in this paper, which are also widely used in many related papers. For these four metrics, larger value means better fusion result. Default parameters provided by respective papers are used and details of these quality metrics are introduced as follows.

1) Normalized Mutual Information (\(Q_{NMI}\)): \(Q_{NMI}\) is an information theory based metric which measures the amount of original information in source images that is maintained in the fused image. Hossny \(\text{et al.}\) [46] modified the unstable traditional mutual information metric [45] to a normalized mutual information metric which is defined as

\[
Q_{NMI} = \frac{\sum_{x=1}^{N_1} \sum_{y=1}^{N_2} \left( \phi^A(x, y) \Phi^F(x, y) + \phi^B(x, y) \Phi^F(x, y) \right)}{\sum_{x=1}^{N_1} \sum_{y=1}^{N_2} \left( \phi^A(x, y) + \phi^B(x, y) \right)},
\]

where \(N_1\) and \(N_2\) are the width and height of a source image respectively. \(\Phi^A\) is the element-wise product of edge strength \(\Phi^A\) and orientation preservation value \(\Phi^F\) of source image \(A\). \(\Phi^B\) is similarly defined for source image \(B\). \(\phi^A\) and \(\phi^B\) are weighting coefficients that reflect the importance of \(\Phi^A\) and \(\Phi^B\) respectively.

2) Gradient-Based Fusion Performance (\(Q_G\)): This gradient based metric is proposed to evaluate how well the sharpness information in source images is transferred to the fused image [47].

\[
Q_G = \sum_{x=1}^{N_1} \sum_{y=1}^{N_2} \left( \phi^A(x, y) \Phi^F(x, y) + \phi^B(x, y) \Phi^F(x, y) \right),
\]

where \(N_1\) and \(N_2\) are the width and height of a source image respectively. \(\Phi^F\) is the element-wise product of edge strength \(\Phi^A\) and orientation preservation value \(\Phi^B\) of source image \(A\). \(\Phi^B\) is similarly defined for source image \(B\). \(\phi^A\) and \(\phi^B\) are weighting coefficients that reflect the importance of \(\Phi^A\) and \(\Phi^B\) respectively.

3) Yang’s Metric (\(Q_Y\)): The structural similarity (SSIM) [48] is used to derive Yang’s metric, which measures the amount of structural information in source images \(A\) and \(B\) that is preserved in the fused image \(F\) [49].

\[
Q_Y = \left\{ \begin{array}{ll}
\gamma_T \text{SSIM}(A, F(T)) + (1 - \gamma_T) \text{SSIM}(B, F(T)), & \text{if } \text{SSIM}(A, B(T)) \geq 0.75 \\
\max\{\text{SSIM}(A, F(T)), \text{SSIM}(B, F(T))\}, & \text{if } \text{SSIM}(A, B(T)) < 0.75
\end{array} \right.
\]

where SSIM(\(T\)) is the structural similarity in a local window \(T\) and \(\gamma_T = \nu(A(T)) / (\nu(A(T)) + \nu(B(T)))\) is the local weight derived from the local variance \(\nu(A(T))\) and \(\nu(B(T))\) within the window \(T\).

4) Chen-Blum Metric (\(Q_{CB}\)): This metric proposed in Chen and Blum’s work [50] computes how well the contrast features are preserved in fused image. Filtering is implemented to each source image and the fused image in the frequency domain. Then masked contrast maps for these images are calculated by the information preservation value map whose detailed definition can be found in [50]. Metric value is then obtained as follows.

\[
Q_{CB} = \sum_{x=1}^{N_1} \sum_{y=1}^{N_2} \left( \Gamma_A(x, y) P_{AF}(x, y) + \Gamma_B(x, y) P_{BF}(x, y) \right),
\]

where \(N_1\) and \(N_2\) are the width and height of the image, \(\Gamma_A\) and \(\Gamma_B\) are the saliency maps for the source image \(A\) and \(B\), respectively. \(P_{AF}\) and \(P_{BF}\) are the information preservation value maps. These two maps are computed from the masked contrast map.

C. Parameters Setting

When calculating the focus measure in (7), the selection of a large block size \(r\) improves robustness to noise but reduces spatial resolution [51]. A radius \(r = 7\) has been found experimentally to be a good tradeoff in this work. The two threshold values in (8) and (9) are set to be \(FM_{th} = 0.02\) and \(R_{th} = 1.2\). The choice of these two parameters is based on observations that a \(7 \times 7\) block with visible texture should
TABLE I

<table>
<thead>
<tr>
<th>Metrics</th>
<th>NSCT</th>
<th>IM</th>
<th>OF</th>
<th>SR</th>
<th>NSCT-SR</th>
<th>DSIFT</th>
<th>EMAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_{NMI}$</td>
<td>0.9422(0)</td>
<td>1.1507(0)</td>
<td>1.0854(0)</td>
<td>1.0922(0)</td>
<td>0.9694(0)</td>
<td>1.2058(15)</td>
<td>1.2069(21)</td>
</tr>
<tr>
<td>$Q_G$</td>
<td>0.7331(0)</td>
<td>0.7437(1)</td>
<td>0.7489(3)</td>
<td>0.7394(0)</td>
<td>0.7345(0)</td>
<td>0.7530(3)</td>
<td>0.7543(29)</td>
</tr>
<tr>
<td>$Q_Y$</td>
<td>0.9470(0)</td>
<td>0.9668(0)</td>
<td>0.9693(1)</td>
<td>0.9571(0)</td>
<td>0.9496(1)</td>
<td>0.9812(1)</td>
<td>0.9863(33)</td>
</tr>
<tr>
<td>$Q_{CB}$</td>
<td>0.7400(0)</td>
<td>0.7698(0)</td>
<td>0.7741(1)</td>
<td>0.7549(0)</td>
<td>0.7452(0)</td>
<td>0.7924(2)</td>
<td>0.7963(33)</td>
</tr>
</tbody>
</table>

have the focus measure larger than 0.02, and the ratio value $R_{th}$ in (9) should be at least 1.2 if this block in that image is identified better focused than the other images. Experiments were conducted to analyze the sensitivity of $FM_{th}$ and $R_{th}$, and results are plotted in Fig. 8 and Fig. 9 respectively. It can be seen that $FM_{th}$ and $R_{th}$ are stable in the nearby range of our choices 0.02 and 1.2, which are marked by the red spots. The threshold for edge width is set to be $\sigma_{th} = 0.5$, since it is shown in [26] that majority of sharp edges have $\sigma$ smaller than 0.5. It is shown in Fig. 10 that the setting of $\sigma_{th}$ is stable in the nearby range of 0.5. Following the settings of image matting in [36], we set the parameter $\lambda$ in (16) to be a large value 10 to keep $w_k$ and $m_k$ consistent at labelled locations. Since the choice of parameters is based on general properties of focus measure and edge model, this universal setting is independent to image content and provides good fusion results for all sets of multi-focus images.

D. Experimental Results and Discussion

1) Objective Evaluation Results: Objective evaluation results of different methods are shown in Table I, all of which are the average metric values over the whole 36 sets of images. The numbers in parentheses denote the number of image sets that this method surpasses the other methods. It is shown that the proposed EMAM fusion method clearly outperforms the other six methods in most cases in terms of the four different evaluation metrics, except $Q_{NMI}$ in some sets of test images. A brief explanation about this is provided here. $Q_{NMI}$ is a good metric since it estimates the mutual information between source images and fused image. But it does not consider whether the local structure of source images is well preserved. It is also reported in [11] that a very high $Q_{NMI}$ value doesn’t always mean good performance. The $Q_{NMI}$ value tends to be larger when the pixel values of the final fused image are closer to one of the source images regardless of whether it is focused or not. A simple test using the ‘block’ image set was conducted and the results are shown in Table II. The $Q_{NMI}$ value when using one of the source images directly as fusion result is dramatically elevated. The DSIFT method uses...
Fig. 11. Gray source images ‘clock’ and the fusion results obtained by different methods. (a) Source image 1. (b) Source image 2. (c) NSCT. (d) GF. (e) IM. (f) SR. (g) NSCT-SR. (h) DSIFT. (i) EMAM.

TABLE II
SIMPLE TEST OF EVALUATION METRICS ON ‘CLOCK’ USING ONE OF THE SOURCE IMAGES DIRECTLY AS FUSION RESULT

<table>
<thead>
<tr>
<th>Method</th>
<th>$Q_{NM1}$</th>
<th>$Q_G$</th>
<th>$Q_Y$</th>
<th>$Q_{GB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed EMAM method</td>
<td>1.2433</td>
<td>0.7531</td>
<td>0.9824</td>
<td>0.7923</td>
</tr>
<tr>
<td>Source image 1 as fusion result</td>
<td>1.3828</td>
<td>0.7164</td>
<td>0.9816</td>
<td>0.7819</td>
</tr>
<tr>
<td>Source image 2 as fusion result</td>
<td>1.3828</td>
<td>0.5824</td>
<td>0.9809</td>
<td>0.7492</td>
</tr>
</tbody>
</table>

...a particular weight strategy that vast majority of the weight map values will be assigned either 0 or 1, which is favored by $Q_{NM1}$ metric. Thus the $Q_{NM1}$ evaluation results of DSIFT for some test image sets are better than the proposed method. However, this strategy may produce some undesirable artifacts, which will be analyzed in detail in a later section IV-E3.

We reckon that the four evaluation metrics and all groups of images should be considered together to evaluate real fusion performance. Besides the objective evaluation, it is also important to examine the performance of different methods through visual comparison, which is shown in the following part.

2) Visualized Results: The fusion results of four groups of images are demonstrated to show advantages of the proposed method. These four groups are chosen to examine the robustness of the proposed method in various situations, including gray or color images, two or more source images, good or poor registration, and complex scenes.

Figs. 11 illustrates a pair of multi-focus source images with good registration and the fusion results obtained by different fusion methods. To reveal the difference between these results, a close-up view of relevant regions in the red box is presented in the left-top of each sub-picture. The main challenge of this pair of images is the boundary region next to the ‘number 8’ of the clock in the back. The boundary of the front clock is so close to the ‘number 8’ that many existing fusion methods are not able to distinguish foreground part and background part explicitly. As shown in Fig. 11i, the fused image obtained by the proposed method perfectly distinguishes these two parts and preserves the complementary information of them, due to the combination of gradient energy focus measure and edge model in the proposed method. The fusion results produced by other methods as shown in Figs. 11c-11h have some kind of erosion in the same region. Also, the sharp boundary of the foreground in source image 2 cannot be reproduced by other methods as well as the proposed EMAM method does.

Figs. 12a and 12b show two multi-focus images capturing a dynamic scene. The head of the person in foreground moved when this pair of images were captured, which is one main reason for misregistration. Results obtained by different methods are illustrated in Figs. 12c-12i. To allow a close examination of the results, we show the normalized difference image which is defined as:

$$I_{Dn}(x, y) = \frac{I_D(x, y) - I_{min}}{I_{max} - I_{min}},$$

where $I_D = I_F - I_r$ denotes the difference image between the fused image $I_F$ and the source image $I_r$ that is best focused in concerned regions. $I_{max}$ and $I_{min}$ are the maximum and minimum values respectively in all difference images of all methods. That is,

$$I_{max} = \max_{all \text{ methods}} \max_{x, y} I_D(x, y)$$

$$I_{min} = \min_{all \text{ methods}} \min_{x, y} I_D(x, y)$$

From an examination of these results, it is obvious that the proposed method well preserves the sharpness information from corresponding source image and excludes the influence of the other. The results of other fusion methods...
contain artifacts in the misregistration region in the red box. Normalized difference images of this misregistration region are generated using source image 2 as a reference, and a close-up view of them is provided in Fig. 13. The good results of the proposed multi-matting method is due to the emphasis of spatial consistency, which is not fully considered in multi-scale decomposition based methods and sparse representation methods. The results of IM method are presented in Fig. 12e and Fig. 13c, which also have good performance in the misregistration region. However, the artifacts in the upper boundary of the foreground, as shown in the blue circle in Fig. 12e, reveal its limitation compared to the proposed method.

Figs. 14a-14b demonstrate a pair of color images with zooming in and out effect, which is another reason for misregistration. This phenomenon widely exists when changing the focus setting to capture images that focus in different objects. As shown in Fig. 14j, the proposed method outperforms the other fusion methods (see Figs. 14d-14i). The selected region in the red box is a good example, where the other methods are more likely to produce artifacts such as color distortion. Even the state-of-the-art method GF suffers from halo artifact in that region. Normalized grayscale difference images of this region are generated using source image 3 as a reference, and a close-up view of them is provided in Fig. 15. The proposed method avoids the halo artifact and generates better results than the other methods, except for the DSIFT method whose results in this region are comparable to ours.

However, comparison of selected region in the blue circle, as shown in Fig. 14i and Fig. 14j, demonstrates the good performance of the proposed method. This is because the proposed multi-matting model is performed in RGB space, while the DSIFT method is performed in grayscale.

Figs. 16a-16b demonstrate a pair of color images captured at complex scenes, where the foreground is a steel mesh and the background can be seen through the holes. This situation is most challenging for spatial-based methods, since the shape of the foreground is not regular. Undesirable artifacts may appear at some narrow foreground regions, where block-based focus measures cannot perform very well. The selected regions in the red box and blue circle shown in Fig. 16 are good examples, in which the spatial-based methods IM and DSIFT mislabel many pixels and have obvious artifacts. The two source images are with perfect registration, so the multi-scale decomposition based methods and sparse representation methods may not suffer from halo and ringing artifacts as described in previous examples. However, the fusion results of them still looks blurry, especially at the boundary of foreground and background. The blurry artifacts are more obvious in the normalized difference images as shown in Fig. 17, which use source image 2 as a reference. It can be seen that method NSCT, SR, NSCT-SR and GF still have a lot of background residuals in the two selected regions. The proposed method’s results as shown in Fig. 16i and 17g have few artifacts with a much clearer boundary of foreground and background.

Fig. 12. Gray source images and the fusion results obtained by different methods. (a) Source image 1. (b) Source image 2. (c) NSCT. (d) GF. (e) IM. (f) SR. (g) NSCT-SR. (h) DSIFT. (i) EMAM.

Fig. 13. Normalized difference images in the red block between the Source 2 and each of the fused images in Fig. 12. (a) NSCT. (b) GF. (c) IM. (d) SR. (e) NSCT-SR. (f) DSIFT. (g) EMAM.
E. Additional Analysis

1) Combination of Edge Model: To investigate the effect of edge model, an experiment was performed without using edge model while other parts of the proposed method remained the same. The final weight map of the foreground using the complete proposed method is shown in Fig. 18b, and the final weight map without using edge model is shown in Fig. 18c. It can be seen that the result without using edge model tends to erode and dilate in the red box, which will lead to poor performance in this boundary region of the final fusion result. Edge model compensates the shortage of the traditional focus measure in this region, and provides a proper guidance for the following multi-matting procedure. The evaluation results are listed in Table III, which also shows the benefit of using edge model.

2) Robustness to Noise: To examine the robustness to noise of the proposed method, experiments were conducted. The ‘clock’ image pair is used as an example for explanation and
similar results were obtained using other image sets. Different levels of Gaussian noise were added to the two source images and the $i$-th noise level corresponds to noise variance $\sigma^2_n = 0.0003i$. 40 levels of noise were used and source images of level 10, 20 and 40 are demonstrated in Fig. 19.

All the seven methods were used to fuse the source images of each noise level. Since there is no groundtruth fusion results, the 4 quality metrics mentioned in Section IV-B are used for evaluation. The evaluation results are shown in Fig. 20, where each sub-figure demonstrates the change of one quality metric as noise level grows. The red solid lines in these figures represent the proposed method, and the performance in metric $Q_{G}, Q_{Y}$ and $Q_{CB}$ are still better than the other methods as noise level grows. The performance in metric $Q_{NMI}$ is not as good as DSIFT whose weight strategy is favored by $Q_{NMI}$ metric. $Q_{NMI}$ value tends to be large when pixel values of the final fused image are

---

**TABLE III**

<table>
<thead>
<tr>
<th>Method</th>
<th>$Q_{NMI}$</th>
<th>$Q_{G}$</th>
<th>$Q_{Y}$</th>
<th>$Q_{CB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>without edge model</td>
<td>1.2418</td>
<td>0.7519</td>
<td>0.9820</td>
<td>0.7693</td>
</tr>
<tr>
<td>with edge model</td>
<td>1.2433</td>
<td>0.7531</td>
<td>0.9824</td>
<td>0.7923</td>
</tr>
</tbody>
</table>
Fig. 18. (a) Source image which focuses in the foreground. (b) Weight map of the foreground using the proposed method. (c) Weight map of the foreground without using edge model.

Fig. 19. (a) Source image with 10-th level noise. (b) Source image with 20-th level noise. (c) Source image with 40-th level noise.

Fig. 20. (a) $Q_{NMI}$ values of different noise levels. (b) $Q_G$ values of different noise levels. (c) $Q_Y$ values of different noise levels. (d) $Q_{CB}$ values of different noise levels.

close to those of one of the source images regardless of whether it is focused or not. A detailed explanation is given in Section IV-D1. However, the $Q_{NMI}$ value of the proposed method doesn’t drop too much compared to the other methods. In general, the evaluation results well verify the proposed method’s robustness to noise.

3) Different Fusion Rule: Nonlinear fusion rule which directly selects the pixel with the largest weight map value at each location was also realized on the proposed method. The final weight maps after multi-matting are used to make these selections. The ‘clock’ image pair is a good representative of the image sets and result of this image pair using linear and nonlinear fusion rule is shown in Fig. 21 with close-up comparison. It can be seen in Fig. 21a that the nonlinear rule produces some sawtooth artifacts in the red circle and blue box regions. This is because that the weight maps become binary masks under the nonlinear fusion rule, and the pixel values may have a sudden change in some boundary regions and depth continually changing regions. Linear fusion rule used in the proposed method produces more acceptable results as shown in Fig. 21b. The evaluation results of these two fusion rules are also listed in Table IV. The linear fusion rule has a better performance in $Q_G$, $Q_Y$ and $Q_{CB}$. But the $Q_{NMI}$ value of nonlinear fusion rule is much higher, which is also a good evidence that $Q_{NMI}$ favors nonlinear fusion rule as stated in Section IV-D1.

4) Computational Efficiency Analysis: All experiments were performed using Matlab R2013b on a computer equipped with a 3.10 GHz CPU and 8 GB memory. The main computation cost is solving the linear system in (17), which can be decomposed into solving several smaller linear systems as stated in (19). Thus the computation complexity of our algorithm is $kF(N_{xy})$, where $k$ is the number of source images and $F(N_{xy})$ is the computation complexity of solving a $N_{xy} \times N_{xy}$ sparse linear system which is well studied in image matting [36], [52]. The average running time over all sets of test images of different fusion methods are compared in Table V. The time of the proposed method is comparable to the SR and NSCT-SR methods, but larger than the others. Encouragingly, our results are based on a simple code implementation and there are many matting techniques like those in the IM method can be used to further improve the computation efficiency when the size and number of input source images increase.
V. CONCLUSIONS

We have proposed an effective EMAM fusion method that outperforms current state-of-the-art fusion methods under various situations. The good performance is due to edge model and the multi-matting model. The combination of edge model and gradient energy focus measure well preserves the sharpness information in source images and defines the boundary of foreground and background accurately. The multi-matting model explicitly generates focus maps and weight maps efficiently in a single optimization problem, and has good performance in boundary and flat regions. These novel algorithms also make the proposed method robust to noise and misregistration in source images.

APPENDIX

DETAILS OF THE MATRIX COMPUTATION OF MULTI-MATTING

Here we define

\[ w = [w_1^T, w_2^T, \ldots, w_N^T]^T, \]

\[ S = [E, E, \ldots, E]_{1 \times N}, \]

where \( E \) is the identity matrix whose height and width are both equal to the number of pixels in one image,

\[ L = \text{Diag}(L_1, L_2, \ldots, L_N), \]

\[ D = \text{Diag}(D_1, D_2, \ldots, D_N), \]

\[ m = [m_1^T, m_2^T, \ldots, m_N^T]^T. \]

So the linear system in (17) can be expressed as:

\[ \begin{bmatrix} L + \lambda D & S^T \\ S & 0 \end{bmatrix} \begin{bmatrix} w \\ \theta \end{bmatrix} = \begin{bmatrix} \lambda D m \\ e \end{bmatrix}. \]

Here we use the properties of block matrix inversion to solve this linear system. First we calculate the Schur complement of the coefficient matrix.

\[ S_{ch} = -S(L + \lambda D)^{-1}S^T. \]

Then we will have

\[ \theta = -S_{ch}^{-1}S(L + \lambda D)^{-1}\lambda D m + S_{ch}^{-1}e \]

\[ w = (L + \lambda D)^{-1}(\lambda D m - S^T \theta) \]

Since \( L + \lambda D \) is a diagonal block matrix, so the inverse of it \((L + \lambda D)^{-1}\) can be easily expressed using the inverse of each diagonal block.

\[ \text{Diag}((L_1 + \lambda D_1)^{-1}, (L_2 + \lambda D_2)^{-1}, \ldots, (L_N + \lambda D_N)^{-1}) \]

Thus the Schur complement is

\[ S_{ch} = -\sum_{k=1}^{N} (L_k + \lambda D_k)^{-1}. \]

\( w \) can be calculated based on these properties:

\[ w = (L + \lambda D)^{-1}\lambda D m - (L + \lambda D)^{-1}S^T \theta \]

\[ = \begin{bmatrix} (L_1 + \lambda D_1)^{-1}L_1 D_1 m_1 \\ (L_2 + \lambda D_2)^{-1}L_2 D_2 m_2 \\ \vdots \\ (L_N + \lambda D_N)^{-1}L_N D_N m_N \end{bmatrix} + \begin{bmatrix} (L_1 + \lambda D_1)^{-1}S_{ch}^{-1}e \\ (L_2 + \lambda D_2)^{-1}S_{ch}^{-1}e \\ \vdots \\ (L_N + \lambda D_N)^{-1}S_{ch}^{-1}e \end{bmatrix} \]

\[ + \sum_{k=1}^{N} ((L_k + \lambda D_k)^{-1}\lambda D_k m_k) \]

Thus each component of \( w \) can be obtained by

\[ w_k = ((L_k + \lambda D_k)^{-1}\lambda D_k m_k) \]

\[ + ((L_k + \lambda D_k)^{-1}S_{ch}^{-1}) \sum_{k=1}^{N} ((L_k + \lambda D_k)^{-1}\lambda D_k m_k) \]

\[ - ((L_k + \lambda D_k)^{-1}S_{ch}^{-1}e) \]

\[ = w_{0k} + ((L_k + \lambda D_k)^{-1}S_{ch}^{-1}(w_{0A} - e)). \]

ACKNOWLEDGMENT

The authors wish to thank Professor Zheng Liu for providing their code implementing objective quality metrics, Professor Yu Liu for providing their MST-SR image fusion toolbox and image dataset, Professor Peter Kovesi for providing their Computer Vision and Image Processing toolbox.

REFERENCES


