QUALITY ESTIMATION BASED MULTI-FOCUS IMAGE FUSION

Jingwei GUAN and Wai-kuen CHAM

The Chinese University of Hong Kong, N.T., Hong Kong

ABSTRACT

In this work, a quality estimation based multi-focus image fusion method (QEBIF) is proposed. In this method, the all-in-focus image is generated by pixel-wise summarizing the multi-focus images with their estimated focus levels as weights. Since the visual quality of an image is highly correlated with its focus level, the visual quality is estimated to be the pre-measurements of focus levels. Via the guided filter, the pre-measurements are smoothed to form the final-measurement with edges in the multi-focus images preserved simultaneously. In addition, the confidence map is proposed to measure the reliability of different local regions. Experiments show that QEBIF method outperforms the other fusion methods, and its fusion results can well maintain the detailed information in the multi-focus images without suffering the ringing or blocking artifacts.

***Index Terms*—** Multi-focus image fusion, visual quality, confidence map, guided filter

1. INTRODUCTION

Various photographs are taken every day. An object appears to be out of focus if the light from the object point is not well converged. Such out-of-focus effect would leads to the image blurring and detailed information missing. Take image A and B in Fig. 1 (a) as an example. These multi-focus images are taken from the same scene. The out-of-focus effect leads to the detailed information missing for the left clock in A and the right clock in B. The multi-focus image fusion task [1] aims at generating an all-in-focus image based on the multi-focus images so that both clocks can be in-focus in this example.

Accurately measuring image focus level is very crucial in generating the all-in-focus image. In this work, we try to introduce an extra image processing task, *i.e.* image quality assessment (IQA), to help do better focus measurement. The IQA task aims at accurately estimating image visual quality, where an out-of-focus image is always estimated of bad quality. Therefore, the visual quality of a image is highly correlated with its focus level. For example in Fig. 1 (b), the focus levels of the images decrease from left to right. Their visual quality also decreases from left to right, which can be easily concluded from the digits in the white patches. Benefiting from this connection, visual quality can be utilized to coarsely measure image focus levels.

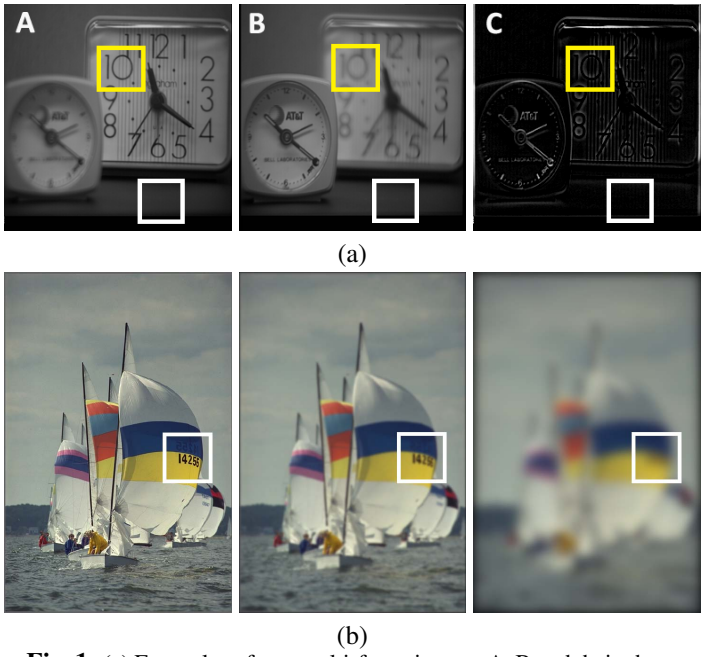


Fig. 1. (a) Examples of two multi-focus images A, B and their absolute difference C. (b) Examples of blur images in the image quality assessment (IQA) dataset. (Best viewed in color.)

To measure the focus levels more accurately, we further analyze the mutli-focus images. These images are supposed to be focused at different positions, and one region cannot be focused in all multi-focus images. However, it is possible for some regions, especially the background, to be out of focus in all multi-focus images. For example in Fig. 1 (a), image C illustrates the absolute difference between A and B, where the brighter color corresponds to the larger difference. For the yellow patches, the difference between A and B is big. Thus the clearer patch in A is very likely to be in-focus and contains more information. While for the white patches, their difference is small. Thus this region is likely to be out-of-focus in both A and B and less informative. In this work, a confidence map is proposed to measure the informative level of different regions. Benefiting from the confidence map, different regions can be treated differently and accordingly.

To sum up, a novel quality estimation based multi-focus image fusion method (QEBIF) is proposed in this work. The contributions are mainly two-fold. First, the visual quality is adopted to help estimate image focus levels benefiting from the similarity between the IQA task and the fusion task. In

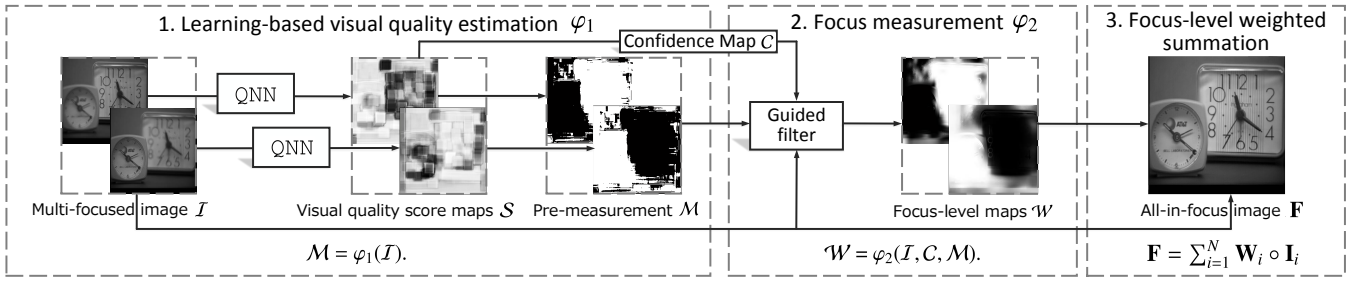


Fig. 2. Pipeline of the proposed quality estimation based multi-focus image fusion method (QEBIF).

addition, the rich images with subjective evaluation results in IQA datasets are able to be utilized to help multi-focus image fusion. Second, the confidence map is explored during the focus measurement. A higher confidence corresponds to a more reliable region. The proposed method employs this concept in the focus measurement and effectively improves the performance.

2. RELATED WORKS

Various multi-focus image fusion methods were proposed in the last decades. The simplest method is averaging (AVG) where the multi-focus images are averaged to generate the all-in-focus image. Generally speaking, the image fusion methods can be divided into three categories, transform-based methods, defocus-modelling methods and spatial-frequency methods. The *transform-based* methods [2, 3, 4, 5, 6, 7, 8] employed various transformations to do the fusion. Liu *et al.* [4] combines MST and SR in the proposed method to overcome the shortcomings of MST- and SR-based image fusion methods simultaneously. Some *transform-based* methods concentrated on different data representations and represented the image in multi-scales [2, 3, 4, 7]. Li *et al.* [2] proposed a fusion method which decomposes the input image into two scales and reconstructs the fusion image using the guided filter. The *defocus-modelling* methods [9, 10] defocus the input images by a designed filter to remove the blur effect, while the *spatial-frequency* methods concentrate on measuring the focus levels of the multi-focus images [11, 12, 13, 14, 15, 16]. Some of the *spatial-frequency* methods adopt pulse coupled neural networks (PCNN) and training methods to implement image fusion and achieve good results [15, 17, 18, 19, 20, 21, 22]. However, these methods do not take the relationship between image visual quality and focus levels into consideration. The proposed QEBIF method is also a spatial-frequency method which concentrates on developing effective focus measurements. The visual quality is utilized to help measure image focus levels, which is different from the previous methods.

3. THE PROPOSED QEBIF METHOD

In this work, a quality estimation based multi-focus image fusion method (QEBIF) is proposed to generate an all-in-focus image $\mathbf{F}$ based on N multi-focus images $\mathcal{I} = \{\mathbf{I}_1, \ldots, \mathbf{I}_N\}$, where $\mathbf{I}_i$ represents i^{th} gray image. The pipeline of the proposed method is summarized in Fig. 2. The all-in-focus image $\mathbf{F}$ is constructed by cumulatively adding the Hadamard product [23] for each multi-focus image $\mathbf{I}_i$ with its focus-level map $\mathbf{W}_i$. *i.e.*

$$\mathbf{F} = \sum_{i=1}^{N} \mathbf{W}_i \circ \mathbf{I}_i, i \in \{1, 2, \ldots, N\}. \quad (1)$$

$\mathcal{W} = \{\mathbf{W}_1, \ldots, \mathbf{W}_N\}$ record the estimated focus level of every pixel (x, y) in every $\mathbf{I}_i$, which is mainly obtained via φ_1 and φ_2 two steps. In φ_1, the visual quality of multi-focus images is estimated as the score maps $\mathcal{S} = \{\mathbf{S}_1, \ldots, \mathbf{S}_N\}$. Based on the correlation between the visual quality and focus level, the pre-measurements of focus levels, $\mathcal{M} = \{\mathbf{M}_1, \ldots, \mathbf{M}_N\}$, are then calculated using $\mathcal{S}$. To sum up, $\mathcal{M} = \varphi_1(\mathcal{I})$.

However, the pre-measurements $\mathcal{M}$ are not ideal for image fusion since $\mathcal{M}$ contains many unwanted sudden changes. Thus an edge-preserving smoothing filter, the guided filter [24], is then utilized to form the final measurement of focus levels $\mathcal{W}$. During this process, the confidence maps $C = \{\mathbf{C}_1, \ldots, \mathbf{C}_N\}$ are employed to help the measurement. To sum up, $\mathcal{W} = \varphi_2(\mathcal{I}, C, \mathcal{M})$. In the following section, φ_1(Sec. 3.1) and φ_2 (Sec. 3.2) would be introduced in details.

3.1. Learning-based visual quality estimation φ_1

To estimate the visual quality score map $\mathbf{S}_i$ of $\mathbf{I}_i$, a deep neural network QNN is proposed, *i.e.* $\mathbf{S}_i = \mathrm{QNN}(\mathbf{I}_i)$. To avoid the block artifact, each pixel in I_i is evaluated individually. In the implementation, each pixel in the image is normalized first. Then, an image patch of size 32×32 centering at (x, y) utilized as the input of QNN to estimate the visual quality of pixel $\mathbf{I}_i(x, y)$.

The architecture of QNN is illustrated in Fig. 3. There are three layer types, *i.e.* convolutional neural layer (Conv), pooling layer (Pooling) and fully connected layer (FC). In the implementation, Rectified linear unit (ReLU) is adopted as the non-linear transformation function of 'Conv' layer. Both the max and min pooling are utilized in the pooling layer.

As discussed in Sec. 1, a region with better quality is more likely to be in-focus. Therefore, the pre-measurement $\mathbf{M}_i$ is calculated as

$$\mathbf{M}_i(x, y) = \begin{cases} 1, \text{ if } \min(\mathbf{S}_1(x, y), \ldots, \mathbf{S}_N(x, y)) = \mathbf{S}_i(x, y), \\ 0, \text{ otherwise}, \end{cases} \quad (2)$$

where a smaller value in $\mathbf{S}_i$ indicates the better visual quality.

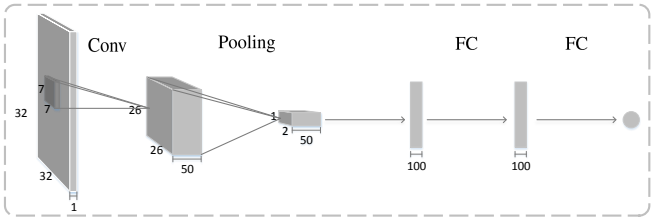


Fig. 3. The architecture of QNN in φ_1.

Thus $\mathbf{M}_i(x, y) = 1$ indicates that $\mathbf{I}_i$ is approximated as the in-focus image at (x, y).

The pre-measurement of focus levels $\mathcal{M}$ can directly be utilized to do the fusion, but the results are unsatisfactory. The reason is that the high non-linearity of QNN network results in the instability of $\mathcal{M}$ as shown in Fig. 2. For example, $\mathcal{M}$ is expected to be smoothed in the clock region, since the whole region is of the same focus level. On the other side, some important lines, such as the boundary between the two clocks, are expected to be well preserved. However, some unwanted sudden changes exists in both cases. The instability of $\mathcal{M}$ directly leads to the instability of the fusion results. Therefore, an edge-preserving smoothing filter, the guided filter [24], is utilized in φ_2 to generate the final measurement of focus levels $\mathcal{W}$.

3.2. Focus measurement process φ_2

For the sake of simplification, $\mathbf{r}$ and $\mathbf{k}$ are utilized to represent pixel (x_r, y_r) and (x_k, y_k) respectively in this section. The edge-preserving smoothing guided filter [24] aims at both smoothing the filter input $\mathcal{M}$ and preserving the lines existing in the guidance image $\mathcal{I}$ simultaneously. To preserve the lines in $\mathcal{I}$, the filter output $\mathcal{W}$ is assumed to be a linear transformation of $\mathcal{I}$. To be specific, $\mathbf{W}_i(\mathbf{k})$ at pixel $\mathbf{k}$ is calculated using $\mathbf{I}_i(\mathbf{k})$ at the same position. For all $\mathbf{k}$ within a square $\omega(\mathbf{r})$ centering at $\mathbf{r}$, $\mathbf{W}_i(\mathbf{k})$ use the same sets of parameters, a($\mathbf{r}$) and b($\mathbf{r}$), to represent the linear relationship, *i.e.*

$$\mathbf{W}_i(\mathbf{k}) = a(\mathbf{r})\mathbf{I}_i(\mathbf{k}) + b(\mathbf{r}), \forall \mathbf{k} \in \omega(\mathbf{r}). \tag{3}$$

The parameters $a(\mathbf{r})$ and $b(\mathbf{r})$ are estimated by minimizing the difference between output $\mathbf{W}_i$ and input $\mathbf{M}_i$ [24]. Equ. (3) is designed to model $\mathbf{W}_i(\mathbf{k})$ from pixel $\mathbf{r}$ only. In fact, pixel $\mathbf{k}$ can be covered by many squares, and all these squares should have an effect on modeling $\mathbf{W}_i(\mathbf{k})$. Therefore, $a(\mathbf{k})$ and $b(\mathbf{k})$ are designed to represent the overall effect of all related $a(\mathbf{r})$ and $b(\mathbf{r})$ on $\mathbf{W}_i(\mathbf{k})$, *i.e.*

$$\mathbf{W}_i(\mathbf{k}) = a(\mathbf{k})\mathbf{I}_i(\mathbf{k}) + b(\mathbf{k}) \tag{4}$$

Generally, a($\mathbf{k}$) and b($\mathbf{k}$) are estimated by averaging all a($\mathbf{r}$) and b($\mathbf{r}$) within $\omega(\mathbf{k})$ [2, 24], *i.e.* $\frac{1}{|\omega|}\sum_{\mathbf{r}\in\omega(\mathbf{k})} a(\mathbf{r})$, $\frac{1}{|\omega|}\sum_{\mathbf{r}\in\omega(\mathbf{k})} b(\mathbf{r})$. $|\omega|$ represents the number of pixels within $\omega(\mathbf{k})$. However, different regions should be treated differently. For example, the orange patch and the gray patch of $\mathbf{I}_1$ in Fig. 4 have different visual quality scores. As discussed in Sec. 1, a region of better quality often has more details and

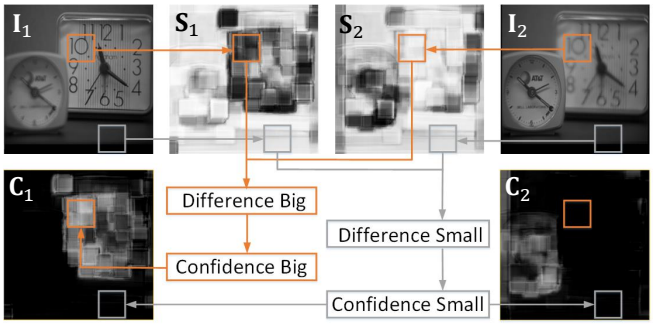


Fig. 4. Illustration of the proposed confidence map **C**. Darker color in **S** represents better quality, while brighter color in **C** represents bigger confidence. (Best viewed in color.)

should be given a higher weight. Thus the confidence maps C are proposed to measure the informative level of different regions.

In this work, C are generated based on the visual quality scores $\mathcal{S}$. If big difference exists among the visual quality of a region in different images, there is a big chance that the region with highest quality is in-focus and contains rich details. On the contrary, things become uncertain for small difference since the region can be out of focus in all images. For example, the orange patch in Fig. 4 shows a bigger difference than the gray patch. The orange patch in $\mathbf{I}_1$ is more informative than the gray one. Thus the orange one can be assigned with a higher weight while the other cannot. To sum up, C can be appoximated by measuring the difference between $\mathcal{S}$, *i.e.*

$$\mathbf{S}_{max}(x, y) = max(\mathbf{S}_1(x, y), \ldots, \mathbf{S}_N(x, y)); \tag{5}$$

$$\mathbf{C}_i(x, y) = max((\mathbf{S}_{max}(x, y) - \mathbf{S}_i(x, y)), T). \tag{6}$$

where $\mathbf{S}_{max}(x, y)$ is the worst visual quality; $T = 0.02$.

With the confidence maps C, a($\mathbf{k}$) are calculated by weighted averaging all $a(\mathbf{r})$ within $\omega(\mathbf{k})$, *i.e.*

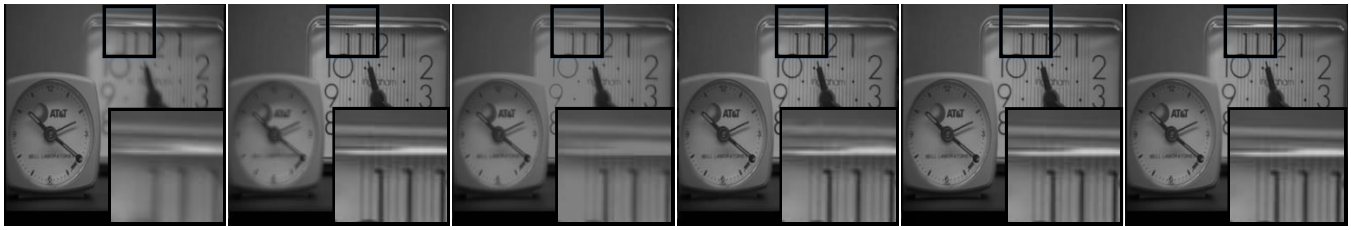
$$a(\mathbf{k}) = \frac{\sum_{\mathbf{r}\in\omega(\mathbf{k})} a(\mathbf{r}) \circ \mathbf{C}_i(\mathbf{r})}{\sum_{\mathbf{r}\in\omega(\mathbf{k})} \mathbf{C}_i(\mathbf{r})}, b(\mathbf{k}) = \frac{\sum_{\mathbf{r}\in\omega(\mathbf{k})} b(\mathbf{r}) \circ \mathbf{C}_i(\mathbf{r})}{\sum_{\mathbf{r}\in\omega(\mathbf{k})} \mathbf{C}_i(\mathbf{r})}. \tag{7}$$

With $a(\mathbf{k})$ and $b(\mathbf{k})$, $\mathbf{W}_i$ can be calculated by Equ. (4)

4. EXPERIMENT

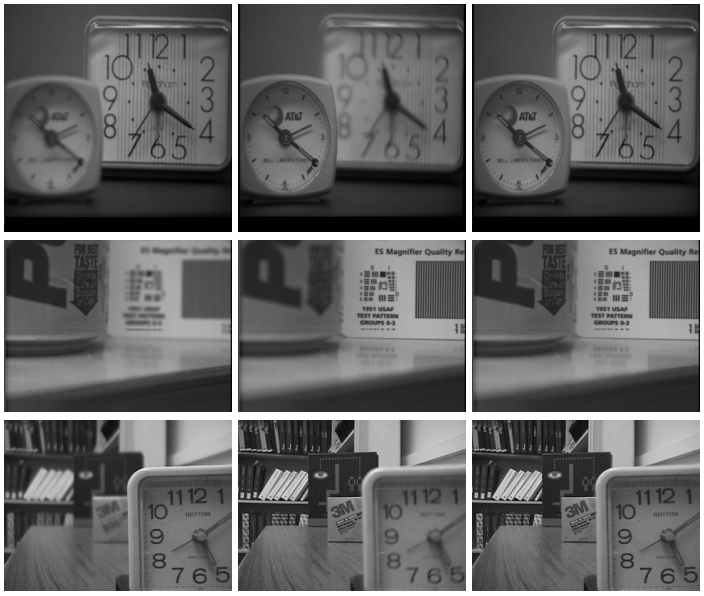
In this section, several experiments were conducted to evaluate the performance of the proposed method. Experimental settings are introduced first. The fusion results of the proposed method are then illustrated and compared with the averaging method (AVG) and some state-of-the-art methods (Sec. 4.1). Finally, component analysis was taken to evaluate the effectiveness of the proposed confidence map (Sec. 4.2).

Experimental Setting: When training QNN, non-overlappingly blur image patches **P** from IQA dataset (the LIVE dataset [25]) are utilized as training samples. Their visual quality scores, difference mean opinion scores (DMOS) provided in the LIVE dataset [25], are utilized as their training labels.



I_1 I_2 AVG Liu *et. al.* [4] Li *et al.* [2] Proposed

Fig. 5. Multi-focus images I_1, I_2 and the fusion results of AVG, Liu *et al.* [4], Li *et al.* [2] and the proposed QEBIF method on 'clock'. The enlarged images of the black patches are shown in the right-bottom corner.



I_1 I_2 F

Fig. 6. The multi-focus images I_1, I_2 and the fusion results F by the proposed QEBIF method. These datasets are 'clock', 'pepsi' and 'disk' from top to bottom.

4.1. Fusion Results

To well explore the performance of the proposed fusion method, three multi-focus image datasets were used for testing, *i.e.* 'clock', 'pepsi' and 'disk'. All images and the fusion results by the proposed method are shown in Fig. 6. It can be seen that the fusion result generated by the proposed QEBIF method are satisfactory, where the details of the multi-focus images are well maintained.

To further explore the performance, AVG and two state-of-the-art fusion methods, Liu *et al.* [4] and Li *et al.* [2], are adopted for comparison as shown in Fig. 5. To see whether the details can be well preserved in the fusion result, the enlarged images of the black patches are also illustrated. For AVG, the information of the multi-focus images is not well preserved in the fusion result, such as the blur digits on the clock. Liu *et al.* [4] suffers from some block artifacts as the observed vertical lines. For Li *et al.* [2], the band outside the boundary is heavier than those in I_2 and the proposed method, which appears to be the ringing artifact. However, the proposed QEBIF method can well preserve the details and does not suffer from the blocking and ringing artifacts.

In addition, the objective evaluation metric MI [26] is utilized to further evaluate these fusion results as summarized in Table 1. Bigger MI indicates better fusion results. It is shown that the proposed QEBIF method achieves the best performance compared with AVG, Li *et al.* [2] and Liu *et al.* [4] on all three datasets.

Table 1. MI [26] on the three datasets.

MI	clock	pepsi	disk
AVG	6.79	5.53	5.39
Liu *et al.* [4]	7.76	7.91	7.62
Li *et al.* [2]	7.79	8.29	7.62
QEBIF	**8.15**	**8.42**	**7.77**

4.2. Confidence map

The confidence map is proposed to help measure the focus levels. In this section, the component analysis was conducted to verify whether the confidence map helps improve the fusion results or not. The fusion results with and without the confidence map are compared by the MI metric [26] as summarized in Table 2. The fusion results with the confidence map achieve better results than those without the confidence map on all datasets, which demonstrates that the proposed confidence map does improve the fusion results.

5. CONCLUSION

In this work, a quality estimation based multi-focus image fusion method (QEBIF) is proposed. The visual quality is estimated to help measure the focus levels since the visual quality of an image is highly correlated with its focus level. In addition, the confidence map is proposed to measure the reliability of different local regions. The fusion results of QEBIF can well maintain the details in the multi-focus images and do not suffer from the ringing or blocking artifacts.

Table 2. Component analysis using MI [26].

MI	clock	pepsi	disk
QEBIF (Without C)	8.06	8.37	7.69
QEBIF (With C)	**8.15**	**8.42**	**7.77**

6. REFERENCES

[1] A Ardeshir Goshtasby and Stavri Nikolov, "Image fusion: advances in the state of the art," *Information Fusion*, vol. 8, no. 2, pp. 114–118, 2007.

[2] Shutao Li, Xudong Kang, and Jianwen Hu, "Image fusion with guided filtering," *IEEE Trans. Image Processing*, vol. 22, no. 7, pp. 2864–2875, 2013.

[3] Gonzalo Pajares and Jesus Manuel De La Cruz, "A wavelet-based image fusion tutorial," *Pattern recognition*, vol. 37, no. 9, pp. 1855–1872, 2004.

[4] Yu Liu, Shuping Liu, and Zengfu Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, pp. 147–164, 2015.

[5] Zhiyu Chen and Shogo Muramatsu, "Multi-focus pixel-based image fusion in dual domain," in *Proc. ICASSP*. IEEE, 2016, pp. 1736–1740.

[6] Zaid Omar, Nikolaos Mitianoudis, and Tania Stathaki, "Region-based image fusion using a combinatory chebyshevica method," in *Proc. ICASSP*. IEEE, 2011, pp. 1213–1216.

[7] Rong Wang, Li-Qun Gao, Shu Yang, Yu-Hua Chai, and Yan-Chun Liu, "An image fusion approach based on segmentation region," *International Journal of Information Technology*, vol. 11, no. 7, pp. 479–486, 2005.

[8] JJ Lewis, RJ Ocallaghan, SG Nikolov, DR Bull, and CN Canagarajah, "Region-based image fusion using complex wavelets," in *International Conference on Information Fusion (FUSION)*, 2004, vol. 1, pp. 555–562.

[9] Muralidhara Subbarao and Tao Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Trans. PAMI*, vol. 17, no. 3, pp. 266–274, 1995.

[10] François Aguet, Dimitri Van De Ville, and Michael Unser, "Model-based 2.5-d deconvolution for extended depth of field in brightfield microscopy," *IEEE Trans. Image Processing*, vol. 17, no. 7, pp. 1144–1153, 2008.

[11] Zhang Yingjie and Ge Liling, "Region-based image fusion using energy estimation," in *Proc. SNPD*. IEEE, 2007, vol. 1, pp. 729–734.

[12] Jing Tian and Li Chen, "Multi-focus image fusion using wavelet-domain statistics," in *Proc. ICIP*. IEEE, 2010, pp. 1205–1208.

[13] Shutao Li and Bin Yang, "Multifocus image fusion using region segmentation and spatial frequency," *Image and Vision Computing*, vol. 26, no. 7, pp. 971–979, 2008.

[14] Said Pertuz, Domenec Puig, Miguel Angel Garcia, and Andrea Fusiello, "Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 1242–1251, 2013.

[15] Qu Xiao-Bo, Yan Jing-Wen, Xiao Hong-Zhi, and Zhu Zi-Qian, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Automatica Sinica*, vol. 34, no. 12, pp. 1508–1514, 2008.

[16] Yibo Chen and Wai-Kuen Cham, "Edge model based fusion of multi-focus images using matting method," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1840–1844.

[17] Shutao Li, James T Kwok, and Yaonan Wang, "Multifocus image fusion using artificial neural networks," *Pattern Recognition Letters*, vol. 23, no. 8, pp. 985–997, 2002.

[18] Zhaobin Wang, Yide Ma, and Jason Gu, "Multi-focus image fusion using pcnn," *Pattern Recognition*, vol. 43, no. 6, pp. 2003–2016, 2010.

[19] Min Li, Wei Cai, and Zheng Tan, "A region-based multi-sensor image fusion scheme using pulse-coupled neural network," *Pattern Recognition Letters*, vol. 27, no. 16, pp. 1948–1956, 2006.

[20] D Agrawal and J Singhai, "Multifocus image fusion using modified pulse coupled neural network for improved image quality," *IET image processing*, vol. 4, no. 6, pp. 443–451, 2010.

[21] Rui Gao, Sergiy A Vorobyov, and Hong Zhao, "Multi-focus image fusion via coupled dictionary training," in *Proc. ICASSP*. IEEE, 2016, pp. 1666–1670.

[22] Yong Yang, Wenjuan Zheng, and Shuying Huang, "Effective multifocus image fusion based on hvs and bp neural network," *The Scientific World Journal*, vol. 2014, 2014.

[23] Roger A Horn, "The hadamard product," in *Proc. Symp. Appl. Math*, 1990, vol. 40, pp. 87–169.

[24] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *Proc. ECCV*. Springer, 2010, pp. 1–14.

[25] Hamid Rahim Sheikh, Muhammad Farooq Sabir, and Alan Conrad Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[26] Guihong Qu, Dali Zhang, and Pingfan Yan, "Information measure for performance of image fusion," *Electronics letters*, vol. 38, no. 7, pp. 1, 2002.