

Local Saliency Extraction for Fusion of Visible and Infrared Images

Weiping Hua, Jufeng Zhao^(✉), Guangmang Cui, Xiaoli Gong, and Liyao Zhu

School of Electronics and Information, Hangzhou Dianzi University,
Hangzhou 310018, China
dabaozjf@hdu.edu.cn

Abstract. In this paper, a local saliency extraction-based dual-band image fusion algorithm is proposed. Combining the variable computational windows, the local gray distance is designed for saliency analysis. And saliency map is further obtained by considering spatial weight. For dual-band image fusion, firstly, we design several local windows named different levels, and get the corresponding saliency maps. Secondly, achieve weighted fusion under different levels with saliency maps. Finally, all fused images are compounded into one fused result. According to experimental results, the proposed method could produce a fused image with good visual effect, preserving even enhancing the details effectively. Comparing with other seven methods, both subjective evaluation and objective metric indicate that the proposed algorithm performs best.

Keywords: Image fusion · Visible and infrared · Local saliency

1 Introduction

Image fusion aims to combine salient information from source images. There exist abundant object details in Visible (VI) image, while infrared (IR) one has particular target characteristic. Fusion for VI and IR images expect to maintain the both advantages of VI and IR images, which is useful for target detection, monitoring, etc.

Lots of methods are designed for image fusion. The fusion based wavelet [1] and curvelet transform are famous multiple resolution-based algorithms. Meanwhile, those pyramid-based approaches also play an important role in image fusion, such as Laplace pyramid [2], ratio pyramid [3], morphological pyramid [4]. Those methods would smooth details because of their down sampling and up sampling, which is time consuming. To achieve image fusion, multiscale directional nonlocal means(MDNLM) filter is used [5]. MDNLM is a multiscale, multidirectional, and shift-invariant image decomposition method. There is an algorithm based on compressive sensing [6], and the sparse coefficients of the source images are obtained by discrete wavelet transform. Bai propose an outstanding algorithm for IR and VI image fusion [7], which utilizes region detection through multi scale center-surround top-hat transform. But some parameters are difficult

to select for new users. To keep the thermal radiation and appearance information simultaneously, people design a fusion algorithm, named Gradient Transfer Fusion (GTF) [8], based on gradient transfer and total variation (TV) minimization. Since human visual system (HVS) is the best system for judging quality of image, saliency preserving is also popular in image fusion [9]. And we also have developed dual-band image fusion using saliency analysis [10, 11]. With saliency characteristic highlighting, the fused results have good visual effect, preserving details well. But how to well extract salient features seems a little difficult for VI and IR image fusion [12]. The saliency extraction method should be designed.

In our paper, we propose a local saliency extraction-based dual-band image fusion method. The local-window-based gray distance idea is used for saliency analysis. And spatial weight is also imposed on the design. The dual-band images are weighted fused based on saliency maps for each band. Furthermore, the fusion is considered under different levels, which are determined by the size of local window to extract features with different sizes. We could finally obtain details enhanced results, inheriting the important information from source images.

2 Local Window-Based Visual Saliency Extraction

HVS could rapidly finish saliency extraction from the scene, as eyes would focus on those areas they concern. Different regions attract different attention, which means HVS would give different weight to the regions. We desire to design algorithms to simulate this ability of HVS to improve the effect of image processing.

Since HVS is sensitive to contrast in visual signal, the histogram-based contrast (HC) method could extract salient object well using color statistics of the input image [13, 14]. In that paper, they operated algorithm in $L^*a^*b^*$ space. This kind of saliency map V is useful for our fusion. V means the weight distribution that HVS pays attention to original image f . And $V \in [0, 1]$, the larger value in V the more attention HVS pays to.

In VI and IR image fusion, we need to highlight characteristic of different sizes, especially small size. Current saliency extraction method with global idea would be out of action. Here, we intend to design a local-based approach to solve this problem. Inspired by the work of Chen [13, 14], we expect to obtain saliency value within a local window with size of $W \times W$. When the value of W changes, the saliency value slightly varies.

At arbitrary pixel (i, j) , the corresponding saliency value $V(i, j)$ for image f is defined as:

$$V(i, j) = \sum_{(i,j) = \Omega_C, \forall (x,y) \in \Omega} [\Gamma(f_{ij}, f_{xy}), D_g(f_{ij}, f_{xy})]. \quad (1)$$

where Ω denotes a local window with size of $W \times W$, whose center is Ω_C located at pixel pixel (i, j) . And (x, y) represents arbitrary pixel in Ω . f_{xy} and f_{ij} are gray value at (x, y) and (i, j) , respectively. In Eq. (1), $D_g(\bullet)$ is gray distance function, which measures the gray distance between the two pixel:

$$D_g(f_{ij}, f_{xy}) = |f_{ij} - f_{xy}|. \quad (2)$$

And $\Gamma(\bullet)$ denotes spatial weight. Here we design this weight because we expect to show the visual difference when the spatial distance changes. $\Gamma(\bullet)$ is defined by the following formula:

$$\Gamma(ij, xy) = e^{\frac{-D_s(ij,xy)}{\sigma^2}}. \tag{3}$$

where $D_s(\bullet)$ means the spatial distance between two pixel using numbers of pixels. And σ^2 is dilatation factor. A large σ^2 makes the distant pixels (x, y) impose large influence on current pixel (i, j) . From the equation, we could find that the closer between (x, y) and (i, j) , the larger influence f_{xy} affect f_{ij} .

The saliency map V is calculated using Eqs.(1), (2) and (3). With this design, we could get salient areas with changing of W . With different levels of $\{W_k\}$, we could obtain corresponding saliency map $\{V_k\}$, and $k = 1, 2, \dots, N$. This could help extracting different features into fused images.

The local computation is operated within a $W \times W$ window as shown in Fig. 1. The saliency value at (i, j) is $V(i, j)$, which is calculated as follows:

- (1) Extract a $W \times W$ image patch P_{ij} , whose center pixel is (i, j) . P_{ij} is segmented from original image f , forming neighborhood Ω .
- (2) Saliency value of center pixel is obtained using P_{ij} . And this saliency value is treated as $V(i, j)$.

Therefore, $V(i, j)$ is computed by following the above two steps. Then, moving this local window pixel-by-pixel, the whole saliency map V could be achieved. Proper size W would emphasize important object in image. And W is usually odd, which could easily determine center pixel. Finally, the function for V is rewritten as $g(\bullet)$:

$$V = g(W, f, \sigma^2). \tag{4}$$

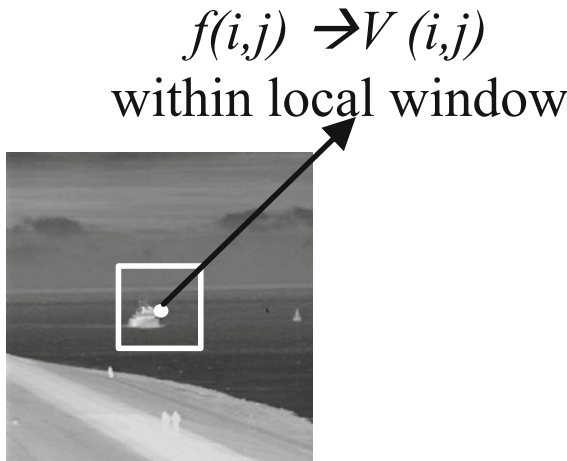


Fig. 1. A local $W \times W$ window for pixel (i, j) to calculate $V(i, j)$

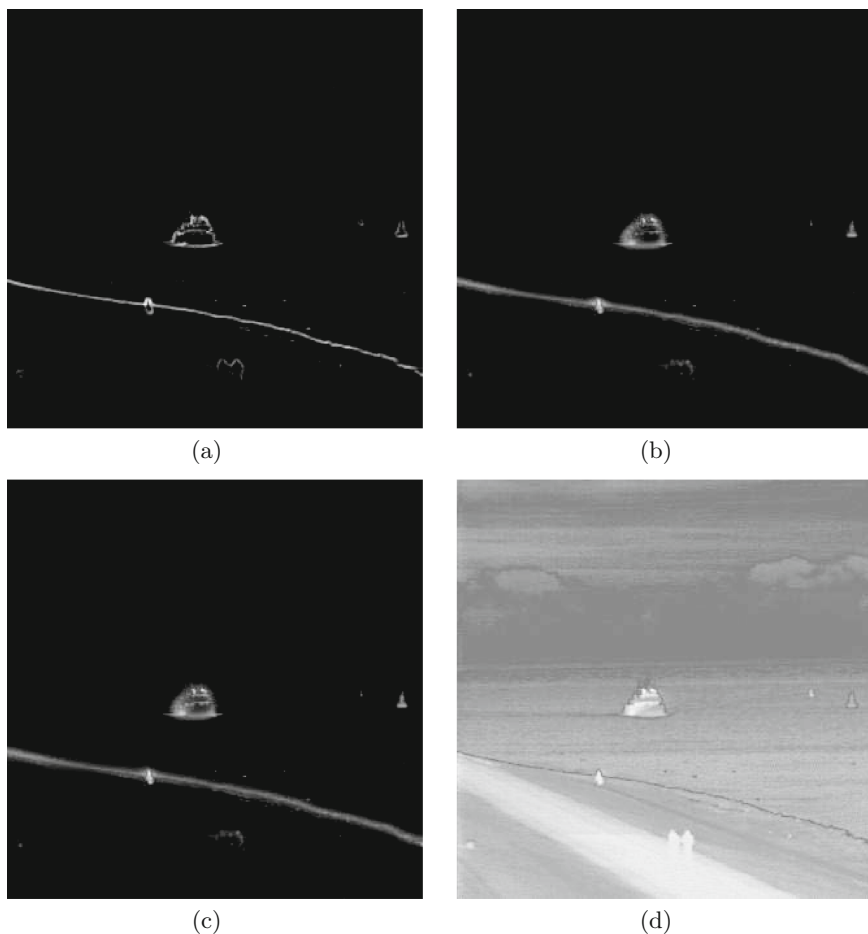


Fig. 2. Saliency maps with local idea and global method: (a) is original image, (b)–(c) are saliency maps with local window ($W=5, 19, \sigma^2 = 3.5$), (d) saliency map using global method.

Figure 2 shows an example of saliency maps, which are generated by our method and global idea [13]. (a) is original infrared image. (b) and (c) are maps created using Eq. (4) with $W = 5$ and 19 , respectively. (d) is the result of global method. We could conclude that, we would extract those characteristic with different sizes, which is very useful in next image fusion.

3 Fusion for Dual-Band Images

According to local-based saliency analysis, the saliency map V could be obtained. Those areas and pixels with large saliency values are expected to be highlighted

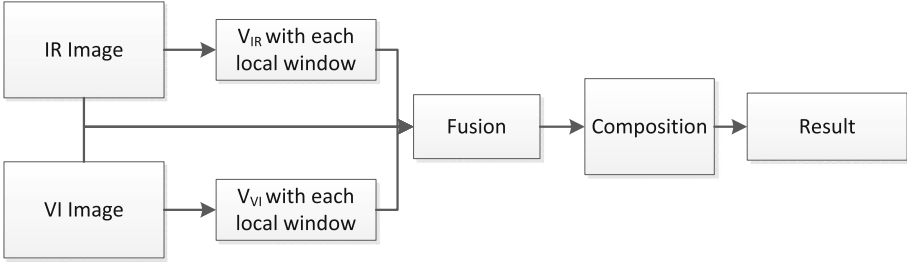


Fig. 3. Flowchart for our fusion

in fused result. We try to utilize multi-window to generate multi-saliency maps, to enhance the details and characteristic have different sizes.

The whole flowchart is shown in Fig. 3. For IR and VI image, the saliency map V_{IR} and V_{VI} are computed under different size levels. And they were fused at each level. The last fused result is achieved by combining those fused image of different levels with proper rules.

With different levels of $\{W_k\}$, we could obtain corresponding saliency map $\{V_k\}$, and $k = 1, 2, \dots, N$. N is the number of size we selected. And $W_{k-1} < W_k$. Following Eq. (4), the k th multiple saliency maps and are created for IR and VI images f_{IR} and f_{VI} , respectively:

$$V_k^{IR} = g(W_k, f^{IR}, \sigma^2). \tag{5}$$

$$V_k^{VI} = g(W_k, f^{VI}, \sigma^2). \tag{6}$$

To enhance the details of IR and VI, respectively, our fusion rule for k th local window is calculated as:

$$F_k = \frac{1}{2} \{ [f^{IR}V_k^{IR} + f^{VI}(1 - V_k^{IR})] + [f^{IR}(1 - V_k^{VI}) + f^{VI}V_k^{VI}] \}. \tag{7}$$

Through above rules, the details are enhanced. Meanwhile, the energy of original images could be preserved.

Finally, we expect to combine those fused images using Eq. (7). In our paper, the fused results need high contrast. Thus, we should emphasize characteristic as more as possible. The final image composition is operated as follows:

$$F = \max_{k=1,2,\dots,N} \{F_k\}. \tag{8}$$

With Eq. (8), the largest value in each pixel of fused images under different levels would be preserved to create a high contrast result. This rule could help produce a details enhanced fused result with good visual effect.

4 Experiment and Discussion

We adopt two image pairs for experiment. The two are downloaded from the weblink: <http://www.google.com>. The images are shown in Fig. 4, which named

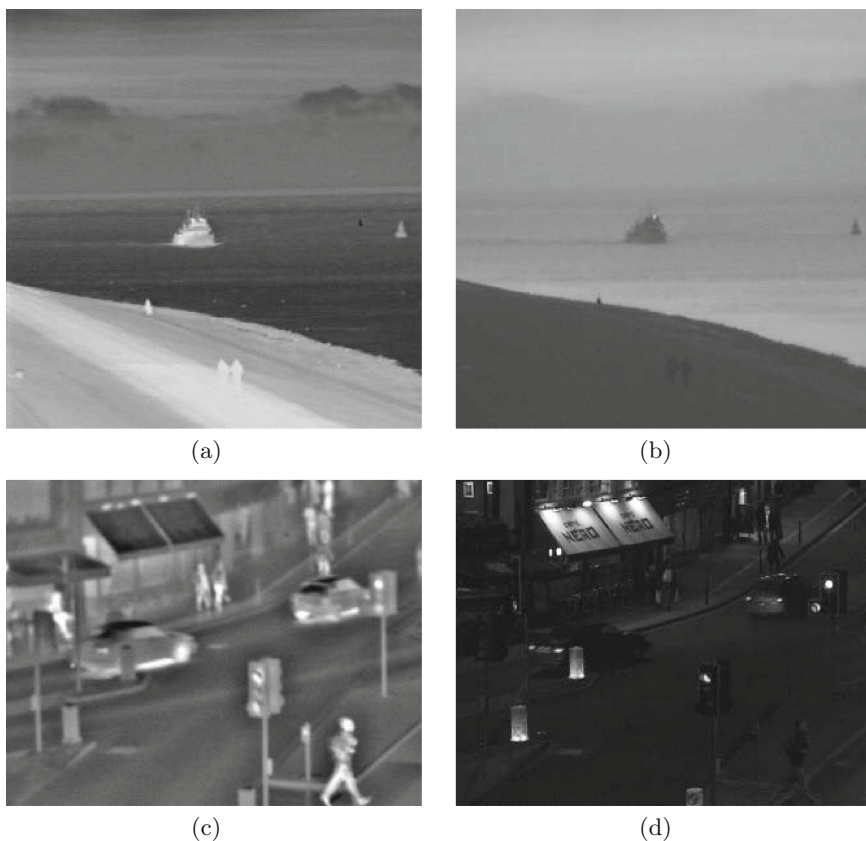


Fig. 4. Source images named ‘boat’ (320×320) and ‘road’ (320×250): (a) and (b) are IR and VI images named ‘boat’, (c) and (d) are IR and VI images named ‘road’.

‘boat’ and ‘road’. (a) and (b) are IR and VI images named ‘boat’. While the other two are corresponding images named ‘road’.

To prove the validness of the proposed method, seven fusion approaches are introduced for comparison. Those algorithms includes direct average algorithm (Direct), wavelet-based algorithm (Wavelet) [1], Laplacian pyramid method (LapP) [2], ratio pyramid approach (RatioP) [3], morphological pyramid (MorP) [4], multi scale center-surround top-hat transform based algorithm (MSCT) [7], saliency preserving method (SalPr) [9].

In order to do further analysis, more data is considered. The other four famous image databases that used in Ref. [10] are adopted. These four datasets are named “UNcamp”, “Dune”, “Trees” and “Octec”, respectively. This four databases are downloaded from the websites www.imagefusion.org in 2012.

All the methods are coded using MATLAB. And the experiments are run on a personal computer (i5-2310 with 2.9 GHz CPU, 4 GB memory).

4.1 Experimental Setting

The parameters W_k and σ^2 in Eqs. (5) and (6) should be determined. The size W_k is expected to cover different salient regions. A large N would make the algorithm run slow, but highlight more objects. According to experimental experience, the number of windows $N = 3$, and $\{W_k\}$ ($k = 1, 2, 3$) = 5, 9, 15. This three window sizes are enough. σ^2 determines the influence of surrounding pixels imposing on center one. We think ten pixels distant would decrease to 0.1 of the neighbor of center pixel. Then $e^{-10/\sigma^2} = 0.1$, we get $\sigma^2 \simeq 4.34$.

4.2 Objective Evaluation for Fusion Image

Besides subjective assessment, we adopt Entropy and Joint Entropy as the objective evaluation for image fusion. X, Y are treated as the two source images. And F is fused result. The fused image quality assessment (FQA) is described as follows.

Since the fused image has combined details of two source images, the fused result should own more information. Thus, Entropy is usually used as FQA [7, 12],

$$En = - \sum_{i=0}^{L-1} p_F(i) \log_2(p_F(i)). \quad (9)$$

In Eq. (9), L denote gray level. Usually $L = 256$ for 8 bit-depth image. This formula shows the probability for gray value i in F . If the En is larger, the fused result seems better.

Joint Entropy is another metric, which show the information of fused image inherited from source ones. The joint entropy is defined by [7, 12],

$$JE_{FXY} = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \sum_{k=0}^{L-1} p_{FXY}(i, j, k) \log_2(p_{FXY}(i, j, k)). \quad (10)$$

where $p_{FXY}(i, j, k)$ is the joint probability. This show the probability that gray values in F is i , in image X is j and gray value in image Y is k , respectively. This metric is similar as Entropy. When JE becomes larger, the fused image is believed to be better.

4.3 Experimental Results and Comparison

We have shown the results of image ‘boat’ in Fig. 5. In this figure, the unique information are two boats, several people, and cloud, etc. (a)–(h) are results of our method and other seven algorithms. The Direct method only takes the average of source images, this leads to the decrease of contrast. The wavelet algorithm makes some artifacts, and the object seems blurred. The result of MorP create a terrible beach, which seems an error. MSCT produce a good result, but the two people is not highlighted enough. Result of LapP, RatioP and

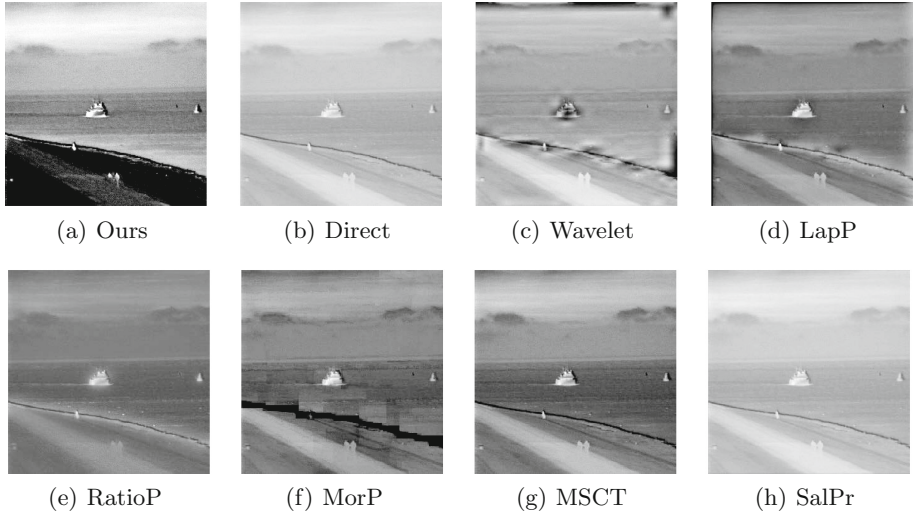


Fig. 5. Fused results of ‘boat’: (a)–(h) are results of proposed method and other seven algorithms, respectively.

Table 1. Quantitative comparison using En, JE for Fig. 5.

FQA/method	Ours	Direct	Wavelet	LapP	RatioP	MorP	MSCT	SalPr
En	7.74	5.67	6.62	6.28	5.72	6.40	6.41	6.31
JE	6.79	6.10	6.42	6.30	6.12	6.34	6.35	6.31

SalPr has low contrast, some characteristic nearly disappeared. The two boats has much higher contrast in our result than other fused ones. In (a), the people, the cloud looks striking. And the beach between sea and land is enhanced. The whole image has better visual effect than other ones.

The result of FQA is listed in Table 1. Our result has the largest value of En and JE, which indicates that our algorithm performs best, inheriting abundant information from the two source images.

The results of ‘road’ are shown in Fig. 6. These two source images own more information than the boat image, such as those people, street lamps, the cars, advertisement words, etc. Observing these results, only our method creates a fused result with high contrast, emphasizing all these features. Table 2 lists the corresponding FQA results. According to judgement of En and JE, the proposed algorithm is outstanding. The largest values of both two metrics demonstrate that our method performs best in dual-band fusion.

4.4 Computational Speed

One of the key metric to evaluate an algorithm is computational speed. In above sections, we have pointed out that the size of boat and road are 320×320 and

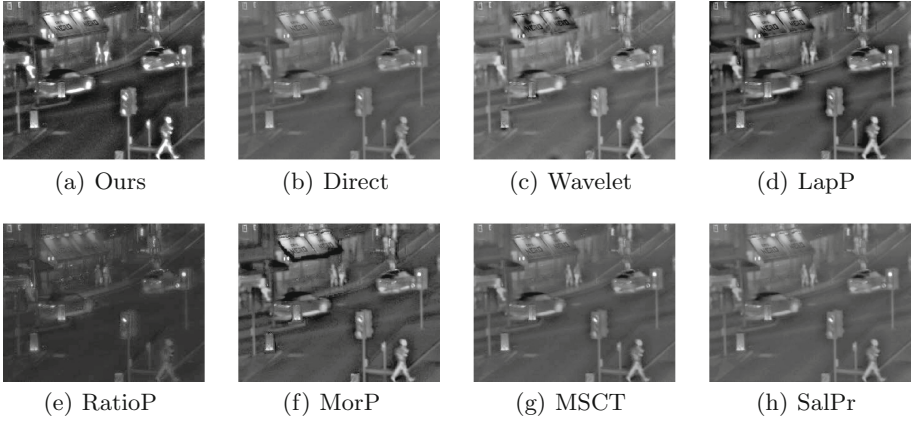


Fig. 6. Fused results of ‘road’: (a)–(h) are results of proposed method and other seven algorithms, respectively.

Table 2. Quantitative comparison using En, JE for Fig. 6.

FQA/method	Ours	Direct	Wavelet	LapP	RatioP	MorP	MSCT	SalPr
En	6.76	5.91	6.40	6.41	4.95	6.30	5.72	5.86
JE	6.05	5.76	5.93	5.93	5.44	5.90	5.71	5.75

Table 3. Comparison of processing time.

Image/methods	Size	Ours	Direct	Wavelet	LapP	RatioP	MorP	MSCT	SalPr
Fig. 5	320×320	1.81	0.02	0.29	0.66	0.03	0.38	3.31	1.62
Fig. 6	320×250	1.62	0.02	0.22	0.52	0.02	0.31	3.09	1.38

320×250 , respectively. All eight algorithms are tested in this section, giving the processing time in Table 3. We could find the fastest method is Direct algorithm. Our method runs slowly because $N = 3$, and the code also need improved and accelerated. If only take one proper local window, the time would be shortened. For example, only use $W = 9$, it only takes 0.4s. When we only use $W = 9$, the fused result is still better than other seven methods. But if we need a much better result, we take $N = 3$ for enhanced. Whatever, in our future work, the algorithm should be further accelerated.

4.5 Visual Performance Discussion

Since HVS is the best criteria, we design subjective evaluation experiment for further visual performance assessment. The detail subjective metric includes “Description of IR image (DIR)”, “Description of VI Image (DVI)”,



Fig. 7. Subjective evaluation device

Table 4. Evaluation results based on eight methods for Fig. 5, using subjective method.

Methods	DIR	DVI	Con	EF
Ours	A	A	A	A
Direct	B	C	C	C
Wavelet	B	C	B	D
LapP	B	C	C	C
RatioP	C	C	C	B
MorP	C	C	B	D
MSCT	B	B	B	A
SalPr	B	B	C	C

Table 5. Evaluation results based on eight methods for Fig. 6, using subjective method.

Methods	DIR	DVI	Con	EF
Ours	A	A	A	A
Direct	B	C	C	C
Wavelet	B	B	C	B
LapP	B	B	B	C
RatioP	C	D	C	C
MorP	B	B	B	A
MSCT	B	B	C	B
SalPr	B	C	C	C

“Contrast(Con)” and “Edge Feature(EF)”. DIR and DVI metrics means the information inherits from the dual-band source images. Con is short for contrast, which is an important metric for fused image. To evaluate details preserving and enhancement, EF is adopted to evaluate sharpness of edges and contours. Four levels are set to describe the quality of images, from worst to the best corresponding to A to D. “D” means very bad, while “C” represents not good, and “B” corresponding to acceptable but not the best, “A” denotes the best level of visual performance.

The subjective results are given by the 20 researchers in our laboratory, the Subjective evaluation device is shown in Fig. 7. And these results are listed in Tables 4 and 5.

Learning from Tables 4 and 5, the proposed method obtains the highest score. Compared with other approaches, the fused result of our approach not only has best global and local contrast, but also sharpest edges. Inspired by subjective evaluation, the proposed method has achieved the best performance in visual effect and details inheritance even enhancement.

5 Conclusion

In this paper, the authors have designed a local saliency extraction-based fusion algorithm for IR and VI images. Utilizing multi-window saliency extraction, the saliency map-based image fusion could create fused image with salient object and information highlighted. Those characteristic with different sizes have been all enhanced. The experiments is proved that this local saliency-based method could well extract feature information of source images, which is effective for dual-band image fusion. The important information will be inherited and enhanced. Both objective assessment and subjective evaluation indicate that the outstanding performance of the proposed method.

It is easy to use our algorithm to achieve dual-band image fusion, enhancing those areas and pixels HVS interests. And the algorithm is suitable for further target detection, scene surveillance and other relative fields, which are our next try.

In future, the local window selection will be automatic on basis of image content. This automation will greatly improve the applicability of algorithm. Meanwhile, how to accelerate the method will be also focused.

Acknowledgments. We thank the reviewers for helping us to improve this paper. Many thanks to Alexander Toet and the TNO Human Factors Research Institute. This work is supported by National Natural Science Foundation of China (Grant No. 61405052). The work is also partly supported by the key technologies R&D Program of Guangzhou city (No. 201704020182). This work is partly supported by State Key Laboratory of Pulp and Paper Engineering (No. 201537), science and technology plan of Zhejiang province (No. 2017C01033) and Key Laboratory Open Fund for RF Circuits and Systems (Hangzhou Dianzi University), Ministry of Education.

References

1. Pajares, G., De La Cruz, J.M.: A wavelet-based image fusion tutorial. *Pattern Recognit.* **37**(9), 1855–1872 (2004)
2. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.* **31**(4), 532–540 (2003)
3. Toet, A.: Image fusion by a ratio of low-pass pyramid. *Pattern Recognit. Lett.* **9**(4), 245–253 (1989)
4. Matsopoulos, G.K., Marshall, S.: Application of morphological pyramids: fusion of MR and CT phantoms. *J. Vis. Commun. Image Represent.* **6**(2), 196–207 (1995)
5. Yan, X., Qin, H., Li, J., Zhou, H., Zong, J., Zeng, Q.: Infrared and visible image fusion using multiscale directional nonlocal means filter. *Appl. Opt.* **54**(13), 4299 (2015)
6. Liu, Z., Yin, H., Fang, B., Chai, Y.: A novel fusion scheme for visible and infrared images based on compressive sensing. *Opt. Commun.* **335**, 168–177 (2015)
7. Bai, X., Zhou, F., Xue, B.: Fusion of infrared and visual images through region extraction by using multi scale center-surround top-hat transform. *Opt. Express* **19**(9), 8444 (2011)
8. Ma, J., Chen, C., Li, C., Huang, J.: Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **31**(C), 100–109 (2016)
9. Hong, R., Wang, C., Wang, M., Sun, F.: Saliency preserving multifocus image fusion with dynamic range compression. *Int. J. Innov. Comput. Inf. Control IJICIC* **5**(8), 2369–2380 (2009)
10. Zhao, J., Gao, X., Chen, Y., Feng, H., Wang, D.: Multi-window visual saliency extraction for fusion of visible and infrared images. *Infrared Phys. Technol.* **76**, 295–302 (2016)
11. Shi, Z., Jiangtao, X., Zhang, Y., Zhao, J., Xin, Q.: Fusion for visible and infrared images using visual weight analysis and bilateral filter-based multi scale decomposition. *Infrared Phys. Technol.* **71**, 363–369 (2015)
12. Guihong, Q., Zhang, D., Yan, P.: Information measure for performance of image fusion. *Electron. Lett.* **38**(7), 313–315 (2002)
13. Cheng, M.-M., Mitra, N.J., Huang, X., Torr, P.H.S., Hu, S.-M.: Global contrast based salient region detection. *IEEE TPAMI* **37**(3), 569–582 (2015)
14. Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: *Computer Vision and Pattern Recognition*, pp. 409–416 (2011)