Fusion of the low-light-level visible and infrared images for night-vision context enhancement

Jin Zhu (朱 进), Weiqi Jin (金伟其)*, Li Li (李 力), Zhenghao Han (韩正昊), and Xia Wang (王 霞)

School of Optoelectronics, Beijing Institute of Technology, Beijing 100081, China
*Corresponding author: jinwq@bit.edu.cn

For better night-vision applications using the low-light-level visible and infrared imaging, a fusion framework for night-vision context enhancement (FNCE) method is proposed. An adaptive brightness stretching method is first proposed for enhancing the visible image. Then, a hybrid multi-scale decomposition with edge-preserving filtering is proposed to decompose the source images. Finally, the fused result is obtained via a combination of the decomposed images in three different rules. Experimental results demonstrate that the FNCE method has better performance on the details (edges), the contrast, the sharpness, and the human visual perception. Therefore, better results for the night-vision context enhancement can be achieved.

OCIS codes: 350.2660, 040.3780, 100.2980, 110.3080.
doi: 10.3788/COL201816.013501.

The low-light-level visible images always provide the details and background scenery, while the target is often detected/recognized via the infrared imaging[1,2] in night vision. As the visible and infrared image fusion technology can improve the perception of the scene in addition to the ability to detect/recognize the target[3], the fusion technology of the low-light-level visible and infrared images plays a significant role in night vision and has been successfully applied in the areas of defense and security[4].

However, night-vision images usually have relatively strong noise, low contrast, and unclear details (including edges). Moreover, human eyes are very sensitive to the details and noise. As these factors have not been considered in most proposed fusion methods, it is difficult to achieve good results in night vision. Thus, an appropriate fusion technology is required for night vision to obtain better results for the night-vision context enhancement.

Liu *et al.* proposed a modified method to fuse the visible and infrared images for night vision[5]. In the method, the visible image is enhanced via the corresponding infrared image, and the fused result is obtained using a conventional multi-scale fusion method. The details of the visible image are not fully enhanced. Salient targets in the infrared image are displayed in dark pixels, which is not good for visual perception. A fusion method for low-light visible and infrared images based on contourlet transform is proposed[6]. Different rules are used for the combination of the low-frequency and high-frequency information. The details of the visible image are not fully enhanced either. Zhou *et al.* proposed a guided-filter-based context enhancement (GFCE) fusion method for night vision[7]. In the result of the GFCE method, the noise has been amplified along with the detail enhancement, and some distortions may emerge in the bright regions due to over enhancement. In all of these methods, neither a denoising method nor a detail enhancing method is used.

Furthermore, the details (including edges) cannot be preserved well enough during the fusion process. Thus, further research needs to be done to obtain better context-enhancement results for the low-light-level visible and infrared imaging.

In order to address the above problems for better night-vision applications, a low-light-level visible and infrared images fusion framework for night-vision context enhancement (FNCE) is proposed in this Letter, as shown in Fig. 1. Actually, the FNCE method can be divided into two parts: the initial enhancement and the fusion process. In the initial enhancement, an adaptive brightness stretching method has been first proposed to enhance the visibility of the low-light-level visible image. At the same time, the denoising and detail enhancement methods are applied for source images. As the multi-scale
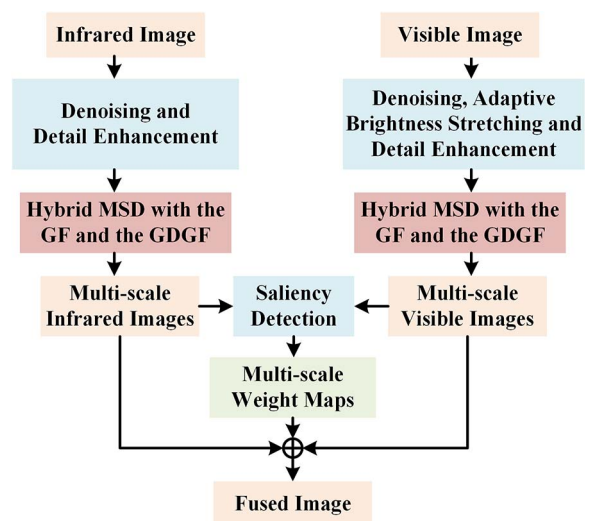

Fig. 1. The proposed infrared and visible image FNCE.

decomposition (MSD) method based on edge-preserving filtering can accurately extract the details at different scales[8]. Furthermore, the gradient domain guided image filtering (GDGF)[9] has better edge performance than the guided image filtering (GF)[10]. Therefore, in the fusion process, a structure of the hybrid MSD with the GF and the GDGF has been proposed to fully decompose the enhanced source images. In addition, the multi-scale weight maps are obtained using a perception-based saliency detection technology at each scale. Finally, the fused result is obtained via the combination of decomposed images with the multi-scale weight maps in three different combination rules according to different scales.

The "Queen's Road," as shown in Fig. 2(a), is collected from the website http://www.imagefusion.org/. The "Buildings" source images, as shown in Fig. 2(b), are captured by a low-light-sensitive CMOS camera and a mid-wave infrared camera. The two test pairs are the typical scenes of urban surveillance applications. As shown in Fig. 2, the source images are usually displayed with unclear details, as well as some noise. Moreover, the contrast of the low-light-level visible image is always low. Thus, as shown in Fig. 1, some enhancement methods must be applied to the source images before the fusion process.

Firstly, a denoising method with the edge-preserving GDGF[9] is used for both the source images to reduce the noise. The values of the filtering parameters of the GDGF are $r = 2$, $\lambda = 0.0001$ for the low-light-level visible image, and $r = 3$, $\lambda = 0.001$ for the infrared image, respectively.

Following this, an adaptive brightness stretching method is proposed for enhancing the visibility of the low-light-level visible image as follows:

$$I_{BS} = \begin{cases} \frac{m}{\mu_I} I, & \text{if } I < \mu_I \\ I, & \text{if } I \geq H \\ \frac{H-m}{H-\mu_I}(I - \mu_I) + m, & \text{if } \mu_I \leq I < H \end{cases}, \quad (1)$$

where I is the input image, I_{BS} is the stretched image, μ_I is the mean of the input image, and (μ_I, m) and (H, H) are the two inflection points of the piecewise linear stretching. As shown in Fig. 3, parts smaller than μ_I will be linearly stretched to m, parts larger than H will be retained, and the rest will be linearly mapped between m and H. The values of parameters in our work are $m = 3\mu_I$, $90 < m < 150$, and $H = 220$. The μ_I is stretched to three times that of the original. At the same time, m should be between 90 and 150, where the mean of the image with
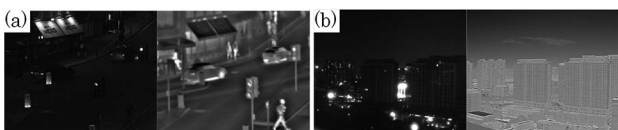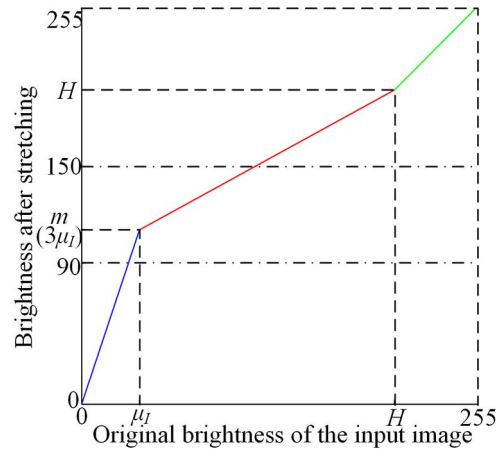

Fig. 3. Adaptive brightness stretching method.

normal illumination is. Therefore, values of the low-light-level visible image can be effectively enhanced to an appropriate range without insufficient or over enhancement via the proposed adaptive brightness stretching method.

Finally, the detail enhancement method with the GF[10] is applied to both source images. The values of the filtering parameters of the GF are $r = 2$, $\lambda = 0.5$, the detail layer is boosted 2.5 times for the low-light-level visible image, and $r = 3$, $\lambda = 0.5$, boosted 3 times for the infrared image.

The initial enhancement results for the "Queen's Road" visible image are shown in Fig. 4. Close-up views for the labeled regions are presented. It can be seen that more and clearer details are presented with less noise using our enhancing method. Thus, the proposed enhancing method for the low-light-level visible image is more effective.

In order to make full use of the information at different scales, a hybrid MSD with the GF and the GDGF is proposed to decompose both source images. The structure of the proposed hybrid MSD is designed as shown in Fig. 5. The GDGF is used to obtain the details of the image (including edges). As an adequate amount of low-frequency information is difficult to obtain via the GDGF, the low-frequency information will be obtained by using the strong GF. As shown in Fig. 5, there are three different levels: the small-scale detail level, the large-scale detail level, and the

Fig. 2. Test pairs of visible and infrared images captured under low-light-level conditions. (a) "Queen's Road," (b) "Buildings."
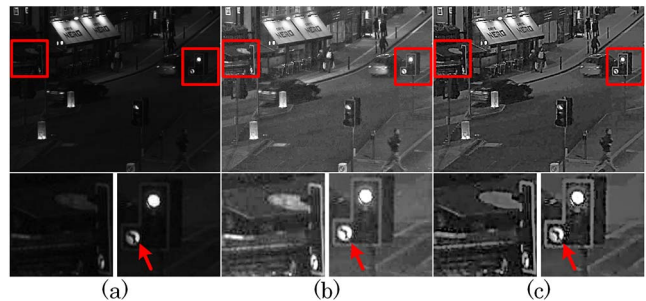

Fig. 4. Initial enhancement results for the "Queen's Road" visible image. (a) The original, (b) result with Zhou's method[7], and (c) result with the proposed method.
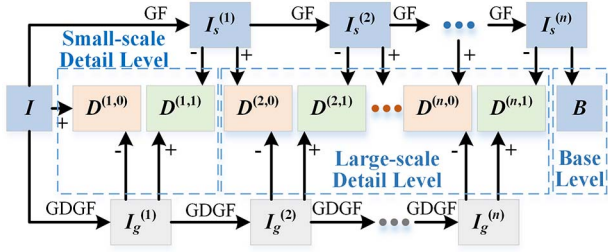
Fig. 5. Structure of the hybrid MSD with the GF and the GDGF.

base level in the decomposition. As the finer details for an image are at the first scale of the hybrid MSD, detail images from the first scale are regarded as the images of the small-scale detail level. The detail images from the second to the nth scale are regarded as the images of the large-scale detail level. The coarsest-scale information is obtained for the base level and can roughly represent the energy distribution.

In the proposed hybrid MSD structure, texture information $D^{(i,0)}$ and edge information $D^{(i,1)}$ at the ith scale are respectively computed as

$$
\begin{aligned}
D^{(i,0)} &= I_s^{(i-1)} - I_g^{(i)}, \\
D^{(i,1)} &= I_g^{(i)} - I_s^{(i)},
\end{aligned} \quad (i = 1, ..., n), \quad (2)
$$

where $I_s^{(i)}$ and $I_g^{(i)}$ are the filtered images at the ith scale with the GF and the GDGF, respectively, both $I_s^{(0)}$ and $I_g^{(0)}$ are the input image I, and n is the number of decomposition scales. $I_s^{(i)}$ and $I_g^{(i)}$ can be obtained as follows:

$$
\begin{aligned}
I_s^{(i)} &= \text{GF}_{r_s^{(i)}, \lambda_s}(I_s^{(i-1)}, I_s^{(i-1)}), \\
I_g^{(i)} &= \text{GDGF}_{r_g^{(i)}, \lambda_g^{(i)}}(I_g^{(i-1)}, I_g^{(i-1)}),
\end{aligned} \quad (i = 1, ..., n), \quad (3)
$$

where $r_s^{(i)}$ and λ_s are the filtering parameters of the GF at the ith scale, $r_s^{(i+1)} = k r_s^{(i)}$, k is a decomposition factor between adjacent scales, λ_s is set to be very large (1×10^4) to acquire low-frequency information; $r_g^{(i)}$ and $\lambda_g^{(i)}$ are the filtering parameters of the GDGF at the ith scale, $r_g^{(i+1)} = k r_g^{(i)}$, and $\lambda_g^{(i+1)} = \lambda_g^{(i)}/k$. Furthermore, as shown in Fig. 5, let the filtered image of the GF at the nth scale $I_s^{(n)}$ be the base image B.

In the fusion process, three different combination rules are respectively designed for the three different levels. The frequency-tuned filtering computes bottom-up saliency[11]. The output of the saliency model strongly correlates with human visual perception[12]. For better weight maps, the frequency-tuned filtering is used to extract the saliency information from the background at each scale. As the targets are always more significant in the infrared images, the weight maps are mainly based on the infrared information, and the infrared characteristic information of the target is maximally highlighted at each scale.

The fused image Fuse is obtained via weighted combinations of the decomposed images as follows:

$$
\text{Fuse} = F_B + \sum_{i=1}^{n} F_D^{(i)}, \quad (4)
$$

where F_B is the base fused image, and $F_D^{(i)}$ is the detail fused image at the ith scale.

For the small-scale detail level, the fused image $F_D^{(1)}$ can be obtained as

$$
F_D^{(1)} = \sum_j [W_{\text{IR}}^{(1,j)} D_{\text{IR}}^{(1,j)} + (1 - W_{\text{IR}}^{(1,j)}) D_{\text{Vis}}^{(1,j)}], \quad (j = 0, 1),
$$
$$(5)$$

where $D_{\text{IR}}^{(1,j)}$ and $D_{\text{Vis}}^{(1,j)}$ are the infrared and visible images at the first scale, and $W_{\text{IR}}^{(1,j)}$ is the weight maps of the infrared images for the small-scale detail level.

The saliency maps of the infrared and visible images for the small-scale detail levels $S_{\text{IR}}^{(1,j)}$ and $S_{\text{Vis}}^{(1,j)}$ $(j = 0, 1)$ are obtained using the frequency-tuned filtering for $D_{\text{IR}}^{(1,j)}$ and $D_{\text{Vis}}^{(1,j)}$ $(j = 0,1)$, respectively. Following this, the binary weight maps of the infrared $\text{BW}_{\text{IR}}^{(1,j)}$ are computed as

$$
\text{BW}_{\text{IR}}^{(1,j)} = \begin{cases} 1 & \text{if } S_{\text{IR}}^{(1,j)} \geq S_{\text{Vis}}^{(1,j)} \\ 0 & \text{otherwise} \end{cases}, \quad (j = 0, 1). \quad (6)
$$

The resulting binary weight maps are noisy and are typically not well aligned with object boundaries. Therefore, spatial consistency is restored through the GDGF with the corresponding detail images $(D_{\text{IR}}^{(1,j)})$ used as guidance images. Finally, the weight maps of the infrared images for the small-scale detail level $W_{\text{IR}}^{(1,j)}$ is obtained as follows:

$$
W_{\text{IR}}^{(1,j)} = \text{GDGF}_{r^{(1)}, \lambda^{(1)}}(D_{\text{IR}}^{(1,j)}, \text{BW}_{\text{IR}}^{(1,j)}), \quad (j = 0, 1), \quad (7)
$$

where the values of filtering parameters are $r^{(1)} = r_g^{(1)}$ and $\lambda^{(1)} = \lambda_g^{(1)}/10$.

For the large-scale detail level, the combination rule is similar to that for the small-scale detail level. The fused image for the large-scale detail level $F_D^{(i)}$ is obtained as

$$
\begin{aligned}
F_D^{(i)} &= \sum [W_{\text{IR}}^{(i)} D_{\text{IR}}^{(i,j)} + (1 - W_{\text{IR}}^{(i)}) D_{\text{Vis}}^{(i,j)}], \\
&(i = 2, ..., n; \; j = 0, 1),
\end{aligned} \quad (8)
$$

where $D_{\text{IR}}^{(i,j)}$ and $D_{\text{Vis}}^{(i,j)}$ are the infrared image and visible images at the ith scale, and $W_{\text{IR}}^{(i)}$ is the weight maps of the infrared images at the corresponding scale.

It should be noted that there is only one weight map $W_{\text{IR}}^{(i)}$ for the images at the ith scale for the large-scale detail level. It is because the two kinds of detail images, $D^{(i,0)}$ and $D^{(i,1)}$ $(i = 2, ..., n)$, always have similar structures. However, the edge detail image $D^{(i,1)}$ has better edge performance than the texture detail image $D^{(i,0)}$. Thus, only $D^{(i,1)}$ is used for the large-scale detail level.

The saliency maps of the infrared and visible images at ith scale for the large-scale detail levels $S_{\text{IR}}^{(i,1)}$ and $S_{\text{Vis}}^{(i,1)}$ $(i = 2, ..., n)$ are also obtained using frequency-tuned filtering for the corresponding images $D_{\text{IR}}^{(i,1)}$ and $D_{\text{Vis}}^{(i,1)}$

$(i = 2, ..., n)$. Following this, the binary weight maps of the infrared $\mathrm{BW}_{\mathrm{IR}}^{(i)}$ $(i = 2, ..., n)$ are computed as

$$\mathrm{BW}_{\mathrm{IR}}^{(i)} = \begin{cases} 1 & \text{if } S_{\mathrm{IR}}^{(i,1)} \geq S_{\mathrm{Vis}}^{(i,1)} \\ 0 & \text{otherwise} \end{cases}, \quad (i = 2, ..., n). \quad (9)$$

The binary weight maps $\mathrm{BW}_{\mathrm{IR}}^{(i)}$ $(i = 2, ..., n)$ are also filtered using the GDGF with the corresponding detail images $D_{\mathrm{IR}}^{(i,1)}$ $(i = 2, ..., n)$ as guidance images. Finally, weight maps of the infrared for the large-scale detail level $W_{\mathrm{IR}}^{(i)}$ $(i = 2, ..., n)$ can be obtained as

$$W_{\mathrm{IR}}^{(i)} = \mathrm{GDGF}_{r^{(i)}, \lambda^{(i)}}(D_{\mathrm{IR}}^{(i,1)}, \mathrm{BW}_{\mathrm{IR}}^{(i)}), \quad (i = 2, ..., n), \quad (10)$$

where $r^{(i)} = r_g^{(i)}$, and $\lambda^{(i)} = \lambda_g^{(i)}/10$ for the GDGF.

For the base level, the fused image F_B is computed as

$$F_B = W_{\mathrm{IR}}^B B_{\mathrm{IR}} + (1 - W_{\mathrm{IR}}^B) B_{\mathrm{Vis}}, \quad (11)$$

where B_{IR} and B_{Vis} are the base images of the infrared and visible, respectively, and W_{IR}^B is the weight map of the infrared image for the base level.

The saliency maps of the infrared and visible images for the base levels S_{IR}^B and S_{Vis}^B are obtained via the frequency-tuned filtering for the corresponding base images. Then, the binary weight map of the infrared $\mathrm{BW}_{\mathrm{IR}}^B$ is computed as

$$\mathrm{BW}_{\mathrm{IR}}^B = \begin{cases} 1 & \text{if } S_{\mathrm{IR}}^B \geq S_{\mathrm{Vis}}^B \\ 0 & \text{otherwise} \end{cases}. \quad (12)$$

The binary weight map $\mathrm{BW}_{\mathrm{IR}}^B$ is smoothed using a Gaussian filter to fit the combination of extremely coarse-scale information. Finally, the weight map for the base level W_{IR}^B is obtained as follows:

$$W_{\mathrm{IR}}^B = g_{\sigma_b}(\mathrm{BW}_{\mathrm{IR}}^B), \quad (13)$$

where $\sigma_b = 2r_s^n$ for the Gaussian filtering $g()$.

In order to test the proposed FNCE method, three state-of-the-art fusion methods are selected for comparison: the guided filtering fusion (GFF) method[13], the gradient transfer fusion (GTF) method[14], and the GFCE method[7]. All the comparative methods are implemented using the public codes, where the parameters are set according to the corresponding Letters. In the FNCE method, the number of decomposition scales $n = 4$, the decomposition factor $k = 2$, and the initial values of the GF are $r_s^{(1)} = 3$, and $\lambda_s = 10^4$, as well as the initial values of the GDGF of $r_g^{(1)} = 2$, and $\lambda_g^{(1)} = 0.05$.

It can be seen from the fusion results of the test images in Fig. 6 that the results of the FNCE method have clearer details (including edges), more salient targets, better
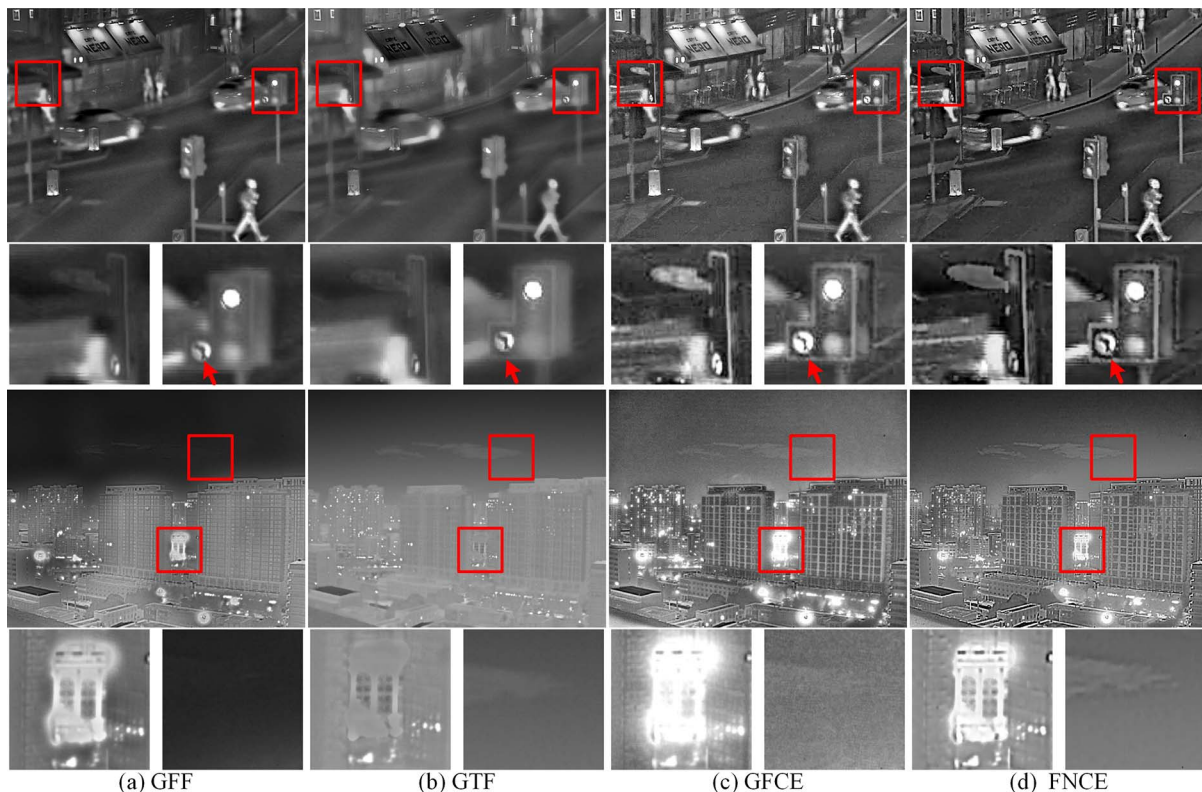

Fig. 6. Fusion results of different methods for the test images.

      (a) GFF   (b) GTF   (c) GFCE   (d) FNCE

contrast, and less noise than other methods. Close-up views for the labeled regions are presented below the images. The results of the GFF method have little detail information from the visible image and are similar than the infrared image with unclear details, as shown in Fig. 6(a). Moreover, the clouds from the infrared image are nearly lost in the "Buildings" result. For the GTF method, as shown in Fig. 6(b), although the results have the least noise, the details are unclear enough. Moreover, lots of information from the visible is lost, for example, the lights. For the GFCE method, as shown in Fig. 6(c), the bright parts (for example, the labeled building with lights) are obviously over-enhanced, and the noise in the sky is obvious in the "Buildings" result. The results of the GFCE method have obvious noise and not clear enough details (edges). Moreover, some distortions may occur due to the over enhancement in the GFCE method. For the FNCE method, as shown in Fig. 6(d), the road sign shown by the red arrow is clearest without distortions in the "Queen's Road" result. Therefore, the proposed FNCE method is able to acquire better results for the human visual perception in night vision.

Information entropy (IE), average gradient (AG), gradient-based fusion metric (Q_G)[15], the metric based on perceptual saliency (PS)[7], and the fusion metric based on visual information fidelity (VIFF)[16] are selected for the objective assessment. IE evaluates the amount of information contained in an image. AG indicates the degree of sharpness. Q_G is recommended for night-vision applications[17] to evaluate the amount of edge information transferred from the source images. PS measures the saliency of perceptual information contained in an image. VIFF evaluates the image quality of the fused image in terms of human visual perception. Table 1 gives the quantitative assessments of different fusion methods on four test image pairs, and the best results are highlighted in bold. The values in Table 1 are averaged values of the four test pairs. It can be seen from Table 1 that IE, AG, PS, and VIFF all achieve the best values in the FNCE method, which means the proposed FNCE method can extract more information, have better sharpness, have more saliency information, and achieve better human visual perception. In addition, the Q_G value of the FNCE method is in the second rank, and it means edges can be relatively better preserved via the FNCE method as well.

The average running time of the different methods on 640×480 source images is shown in Table 2. All of the

Table 1. Quantitative Assessments of Different Methods

Method	IE	AG	Q_G	PS	VIFF
GFF	6.6093	0.0100	**0.6132**	16.4800	0.4278
GTF	6.3275	0.0061	0.2847	13.7586	0.2205
GFCE	6.8106	0.0152	0.3612	19.5798	0.5648
FNCE	**6.9786**	**0.0187**	0.6029	**20.6976**	**0.6580**

Table 2. Average Running Time on 640×480 Images

Method	GFF	GTF	GFCE	FNCE
Time (s)	0.526	6.833	1.605	2.069

compared methods are implemented via MATLAB on a computer (Inter i5 3.40 GHz CPU, 4G RAM).

After the experimental comparisons, it can be seen that better human visual perception is achieved with more salient targets, better details (edges) performance, better contrast, better sharpness, and less noise in the FNCE method. Obviously, the proposed FNCE method is more effective, which will help to obtain better context enhancement for the night-vision imaging. Although the FNCE method is slightly time-consuming, it is acceptable, considering the better fused result.

In conclusion, an FNCE method is proposed. First, an adaptive brightness stretching method is proposed to enhance the visibility of the low-light-level visible image. Following this, a structure of the hybrid MSD with the GF and the GDGF is proposed for fully decomposing the enhanced source images. In addition, weight maps are obtained via a perception-based saliency detection technology at each scale.

Experimental results show that better results for night-vision context enhancement can be acquired via the proposed FNCE method. In the future, the idea of the fast GF[18] may be introduced into the simplifications of the FNCE method for practical applications. Moreover, the previous frame video image may be used as the guidance image of the current frame to reduce the delay.

References
1. Y. Wang, F. Xie, and J. Wang, Chin. Opt. Lett. **14**, 122801 (2016).
2. Z. Qin, G. Xie, J. Ma, P. Yuan, and L. Qian, Chin. Opt. Lett. **15**, 111402 (2017).
3. A. Toet, J. K. IJspeert, A. M. Waxman, and M. Aguilar, Displays **18**, 85 (1997).
4. A. Toet, Opt. Eng. **51**, 010901 (2012).
5. Z. Liu and R. Laganière, Signal Image Video Process. **1**, 293 (2007).
6. S. Liu, Y. Piao, and M. Tahir, Opt. Eng. **55**, 123104 (2016).
7. Z. Zhou, M. Dong, X. Xie, and Z. Gao, Appl. Opt. **55**, 6480 (2016).
8. S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, Inf. Fusion **33**, 100 (2017).
9. F. Kou, W. Chen, C. Wen, and Z. Li, IEEE Trans. Image Process. **24**, 4528 (2015).
10. K. He, J. Sun, and X. Tang, IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1397 (2013).
11. R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE 2009)* (2009), p. 1597.

12. A. Toet, IEEE Trans. Pattern Anal. Mach. Intell. **33**, 2131 (2011).

13. S. Li, X. Kang, and J. Hu, IEEE Trans. Image Process. **22**, 2864 (2013).

14. J. Ma, C. Chen, C. Li, and J. Huang, Inf. Fusion **31**, 100 (2016).

15. C. S. Xydeas and V. Petrovic, Electron. Lett. **36**, 308 (2000).

16. Y. Han, Y. Cai, Y. Cao, and X. Xu, Inf. Fusion **14**, 127 (2013).

17. Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, IEEE Trans. Pattern Anal. Mach. Intell. **34**, 94 (2012).

18. K. He and J. Sun, "Fast guided filter," arXiv: 1505.00996 (2015).