VIFB: A Visible and Infrared Image Fusion Benchmark
Supplementary Document

Xingchen Zhang[1,2], Ping Ye[1], Gang Xiao[1]
[1]School of Aeronautics and Astronautics, Shanghai Jiao Tong University
[2]Department of Electrical and Electronic Engineering, Imperial College London
xingchen.zhang@imperial.ac.uk, {yeping2018, xiaogang}@sjtu.edu.cn

In this supplementary document, we first add introduction to evaluation metrics that have been integrated in the proposed visible and infrared image fusion benchmark (VIFB). Then, we present more qualitative fusion results obtained in VIFB.

1. More evaluation metrics

In VIFB, we have implemented 13 evaluation metrics which are frequently utilized in image fusion. However, due to the page limit, we do not give the definition for each metric in the paper. In this supplementary document, we give the definitions of some evaluation metrics.

In the following definitions, M and N are the width and height of the images, respectively. The subscript V indicates the visible image, the subscript I represents the infrared image, and the subscript F represents the fused image.

1. Cross entropy (CE) [2].

   The CE between the fused image and all source images is defined as:

$$CE = \frac{CE_{V,F} + CE_{I,F}}{2}, \tag{1}$$

   where $CE_{V,F}$ is the cross entropy between visible image and fused image, $CE_{I,F}$ is the cross entropy between infrared image and fused image.

   $CE_{V,F}$ is computed as:

$$CE_{V,F} = \sum_{i=0}^{255} h_V(i) log_2 \frac{h_V(i)}{h_F(i)}, \tag{2}$$

   where $h(i)$ is the normalized histogram of the image.

   $CE_{I,F}$ is computed as:

$$CE_{I,F} = \sum_{i=0}^{255} h_I(i) log_2 \frac{h_I(i)}{h_F(i)}. \tag{3}$$

   A small CE value means the fused image has considerable similarity with source images, thus indicating a good fusion performance.

2. Entropy (EN) [1].

   EN measures the information contained in the fused image. Its definition is:

$$EN = -\sum_{l=0}^{L-1} p_l log_2 p_l, \tag{4}$$

   where L represents the number of gray levels and p_l denotes the normalized histogram of the corresponding gray level in the fused image. A large EN indicates a bettor fusion performance.

3. Mutual information (MI) [10].
   MI is used to measure the amount of information that is transferred from source images to the fused image. It is defined as:
   $$MI = MI_{V,F} + MI_{I,F}, \tag{5}$$
   where $MI_{V,F}$ and $MI_{I,F}$ denote the information transferred from visible and infrared images to the fused image, respectively. Specifically, $MI_{X,F}$ is defined as follows:
   $$MI_{X,F} = \sum_{x,f} p_{X,F}(x,f) log \frac{p_{X,F}(x,f)}{p_X(x)p_F(f)}, \tag{6}$$
   where $p_X(x)$ and $p_F(f)$ are the marginal histograms of source image X and fused image F, respectively. $p_{X,F}(x,f)$ is the joint histogram of source image X and fused image F. A large MI value means a good fusion performance since considerable information is transferred to the fused image.

4. Peak signal-to-noise ratio (PSNR) [7].

   PSNR indicates the ratio of peak value power and noise power in the fused image. It can measure the distortion during the image fusion process and is defined as:
   $$PSNR = 10log_{10} \frac{r^2}{MSE}, \tag{7}$$
   where r is the peak value of the fused image, MSE is the mean squared error computed as:
   $$MSE = \frac{MSE_{V,F} + MSE_{I,F}}{2}, \tag{8}$$
   where $MSE_{V,F} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (V(i,j) - F(i,j))^2$, $MSE_{I,F} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i,j) - F(i,j))^2$. A large PSNR means that the fused image is close to source images and has less distortion. Therefore, the larger the PSNR metric, the better the fusion performance is.

5. Average gradient (AG) [5].

   AG measures the gradient information of the fused image and represents its detail and texture [5]. It is defined as:
   $$AG = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\frac{\nabla F_x^2(i,j) + \nabla F_y^2(i,j)}{2}}, \tag{9}$$
   where $\nabla F_x(i,j) = F(i,j) - F(i+1,j)$, $\nabla F_y(i,j) = F(i,j) - F(i,j+1)$. A large AG value indicates that more gradient information is contained in the fused image and thus means a good fusion performance.

6. Standard devision (SD) [12].

   SD reflects the distribution and contrast of the fused image [9]. Its definition is:
   $$SD = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (F(i,j) - \mu)^2} \tag{10}$$
   where μ represents the mean value of the fused image. The human visual system is sensitive to contrast, thus the regions in an image with high contrast always attract human attention. Since high contrast in a fused image leads to a large SD, thus a large SD indicates that the fused image is with a good visual effect.

7. Spatial frequency (SF) [6].
   SF can measure the gradient distribution of an image thus revealing the detail and texture of an image. It is defined as:
   $$SF = \sqrt{RF^2 + CF^2}, \tag{11}$$
   where $RF = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (F(i,j) - F(i,j-1))^2}$ and $CF = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} (F(i,j) - F(i-1,j))^2}$. A large SF value indicates rich edges and textures, thus indicating good fusion performance.

8. Gradient based similarity measurement ($Q^{AB/F}$) [14].

   $Q^{AB/F}$ indicates the amount of edge information that is transferred form source images to fused image [9]. It can be computed as:

$$Q^{AB/F} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(Q^{V,F}(i,j)w^{V}(i,j) + Q^{I,F}(i,j)w^{B}(i,j))}{\sum_{i=1}^{N}\sum_{j=1}^{N}(w^{A}(i,j) + w^{B}(i,j))}, \tag{12}$$

   where $Q^{X,F}(i,j) = Q_g^{X,F}(i,j)Q_a^{X,F}(i,j)$, $Q_g^{X,F}(i,j)$ and $Q_a^{X,F}(i,j)$ represent the edge strength and orientation values at location (i,j), respectively. w^X denotes the weight that expresses the importance of each source image to the fused image. The more edge information transferred to the fused image, the larger the $Q^{AB/F}$ value is. Thus a large $Q^{AB/F}$ value indicates a good fusion performance.

9. Root mean squared error (RMSE) [7] .
   RMSE is defined as:

$$RMSE = \frac{RMSE_{V,F} + RMSE_{I,F}}{2}, \tag{13}$$

   where $RMSE_{V,F}$ denotes the dissimilarity between the visible and fused images, $RMSE_{I,F}$ is the dissimilarity between the infrared and fused images. $RMSE_{X,F}$ is defined as:

$$RMSE_{X,F} = \sqrt{\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}(X(m,n) - F(m,n))^2}. \tag{14}$$

   If the fused image has a small amount of error and distortion, then there will be a small RMSE value.

10. Structural similarity index measure (SSIM) [13].

    SSIM is used to model image loss and distortion, to which the human visual system is sensitive [13]. It consists of three parts, namely loss of correlation, luminance, and contrast distortion. SSIM for visible or infrared image is defined as the product of these three parts, i.e.

$$SSIM_{X,F} = \sum_{x,f}\frac{2\mu_x\mu_f + C_1}{\mu_x^2 + \mu_f^2 + C_1}\cdot\frac{2\mu_x\mu_f + C_2}{\mu_x^2 + \mu_f^2 + C_2}\cdot\frac{\sigma_{xf} + C_3}{\sigma_x\sigma_f + C_3}, \tag{15}$$

    where $SSIM_{X,F}$ denotes the structural similarity between source image X (X is V for visible image and is I for infrared image) and fused image F, x and f represent the image patches of source and fused image in a sliding window, respectively. σ_{xf} is the covariance of source and fused images, σ_x and σ_f represent the standard deviation, μ_x and μ_f are the mean values of source and fused images, respectively. C_1, C_2, C_3 are the parameters used to make the algorithm stable.

    The structural similarities between the fused image and both source images can be defined as:

$$SSIM = SSIM_{V,F} + SSIM_{I,F}. \tag{16}$$

    A large SSIM value indicates a better fusion performance.

11. Edge intensity (EI) [11].

    EI measures the edge intensity information of an image. A higher EI value indicates more clearness and higher image quality. EI can be computed using Sobel operator as:

$$EI = \sqrt{S_x^2 + S_y^2}, \tag{17}$$

    where

$$S_x = F * h_x, S_y = F * h_y, \tag{18}$$

    where $h_x = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$, $h_y = \begin{bmatrix} 1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$, $*$ is convolution operation.

Regarding other metrics, the definition of Q_{CB} can be found in [4] and [8], the definition of Q_{CV} can be founded in [3] and [8].

2. More qualitative fusion results

Due to the page limit, we just presented the qualitative comparison of two image pairs in the paper. In this supplementary document, we add qualitative results of more image pairs obtained in VIFB.

Figure 1 presents the qualitative results of *carLight* image pair. In this case, GFF, MGFF, amd TIF give relative good fusion performance. In the fused images obtained by these three approaches, the car which is invisible in the visible image due to over-exposure can be seen, and less artifacts are produced.

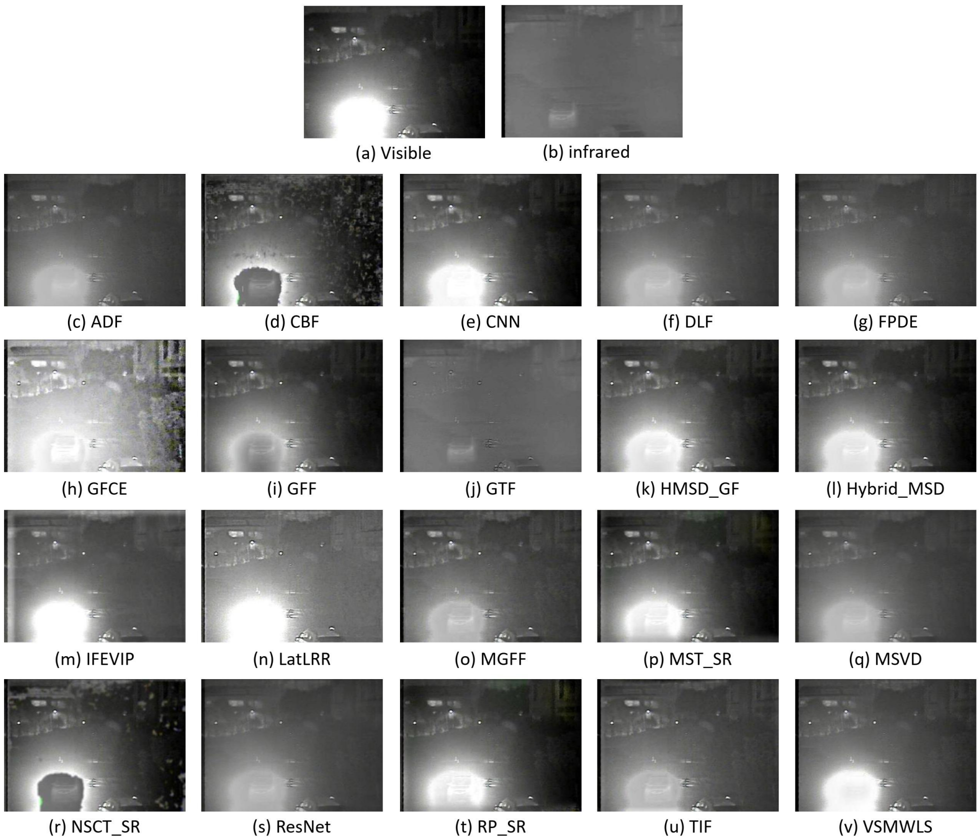

Figure 1. The fusion results of the *carLight* image pair.

Figure 2 shows the results of *kettle* image pair. In this case, IFEVIP, LatLRR, MGFF, and Hybrid_MSD give relative good qualitative results. In the fused images produced by these methods, the man with a hot kettle can be seen clearly. Besides, the details and textures in visible image are well preserved.

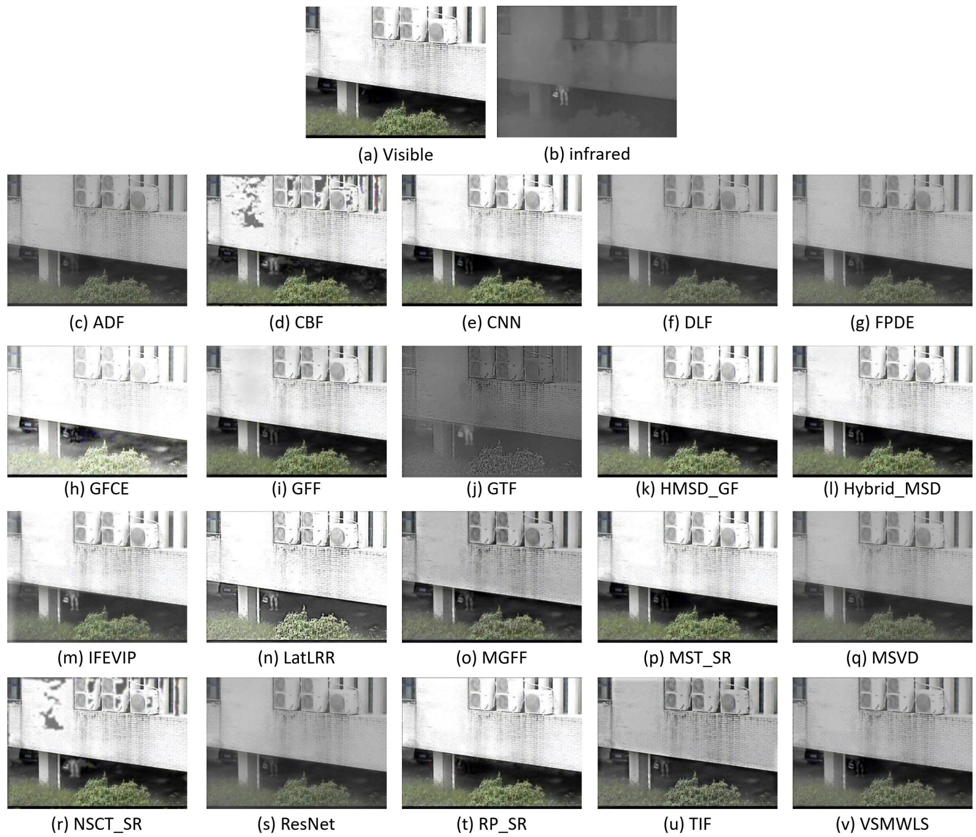

Figure 2. The fusion results of the *kettle* image pair.

Figure 3 illustrates the qualitative results of *carShadow* case. In this case, several algorithms give visually good fused images, namely DLF, GFF, MGFF, MSVD, ResNet, TIF and VSMWLS. However, some algorithms produce obvious artifacts in the fused image, such as CBF and NSCT_SR.

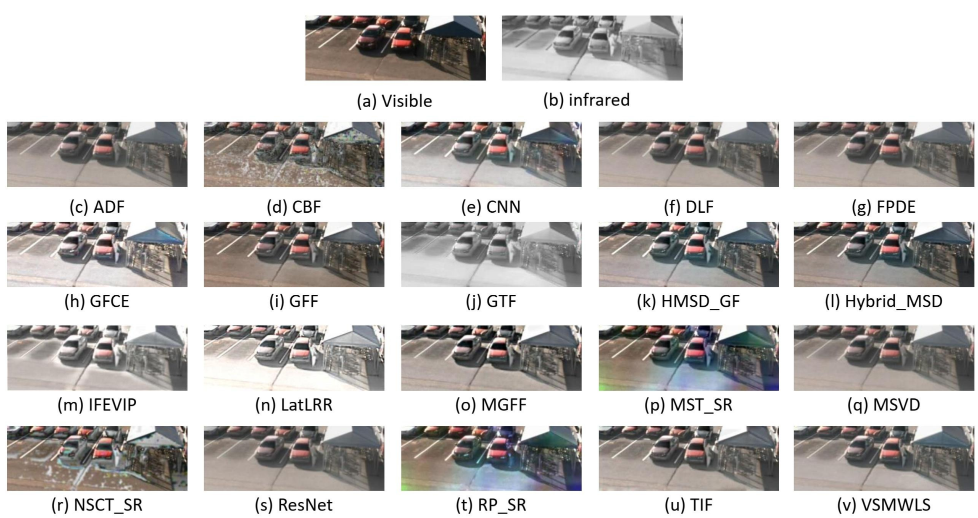

Figure 3. The fusion results of the *carShadow* image pair.

Figure 4 presents fusion results of *manCall* case. As can be seen, GFCE, Hybrid_MSD, LatLRR, TIF and VSMWLS give relative good qualitative results. In the fused images produced by these methods, the man who is making a phone call can be

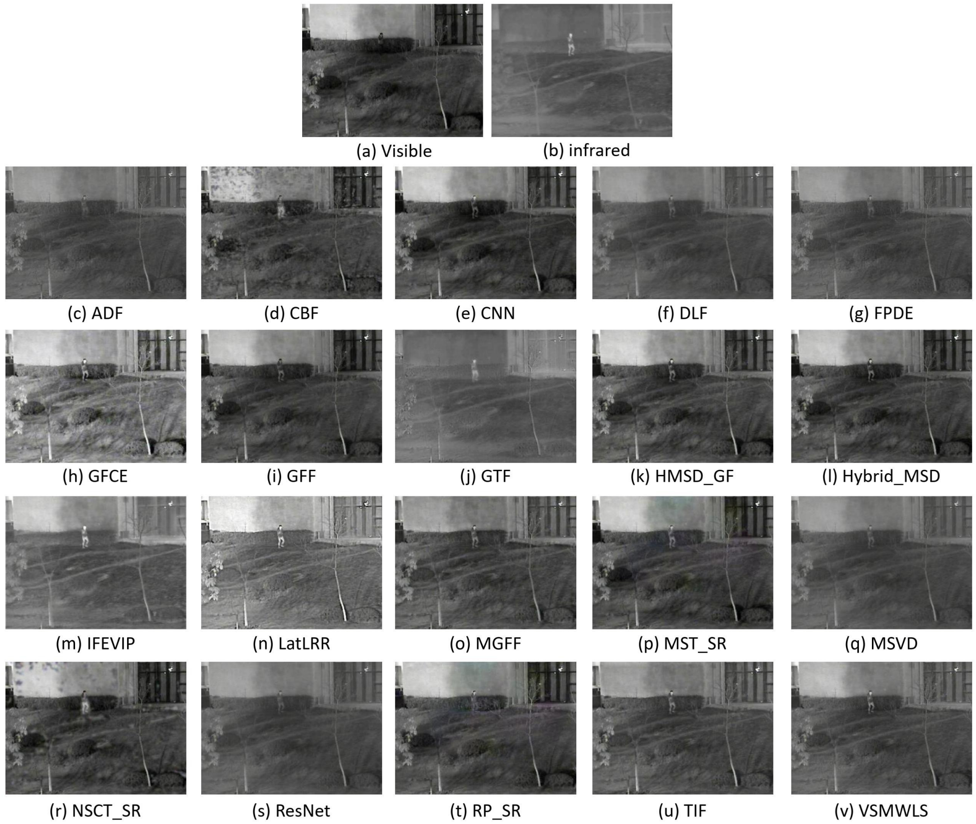

Figure 4. The fusion results of the *manCall* image pair.

seen clearly. Furthermore, the details and textures in visible image, for example the grass and tree, are well preserved in the fused images.

The fusion results presented above indicate that the performance of an image fusion method may vary when applied to different image pairs. At the moment, there is not a method which can obtain the best results on all image pairs. Therefore, it is necessary to evaluate image fusion algorithms using different kinds of methods and metrics.

References

[1] V. Aardt and Jan, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," Journal of Applied Remote Sensing, vol. 2, no. 1, p. 023522, 2008.

[2] D. M. Bulanon, T. Burks, and V. Alchanatis, "Image fusion of visible and thermal images for fruit detection," Biosystems Engineering, vol. 103, no. 1, pp. 12–22, 2009.

[3] H. Chen and P. K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," Information fusion, vol. 8, no. 2, pp. 193–207, 2007.

[4] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," Image and vision computing, vol. 27, no. 10, pp. 1421–1432, 2009.

[5] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," Optics Communications, vol. 341, pp. 199 – 209, 2015.

[6] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," IEEE Transactions on communications, vol. 43, no. 12, pp. 2959–2965, 1995.

[7] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," Aquatic Procedia, vol. 4, no. Icwrcoe, pp. 133–142, 2015.

[8] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 94–109, 2012.

[9] J. Ma, Y. Ma, and C. Li, “Infrared and visible image fusion methods and applications: A survey,” Information Fusion, vol. 45, pp. 153–178, 2019.

[10] G. Qu, D. Zhang, and P. Yan, “Information measure for performance of image fusion,” Electronics letters, vol. 38, no. 7, pp. 313–315, 2002.

[11] B. Rajalingam and R. Priya, “Hybrid multimodality medical image fusion technique for feature enhancement in medical diagnosis,” International Journal of Engineering Science Invention, 2018.

[12] Y.-J. Rao, “In-fibre bragg grating sensors,” Measurement science and technology, vol. 8, no. 4, p. 355, 1997.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli et al., “Image quality assessment: from error visibility to structural similarity,” IEEE transactions on image processing, vol. 13, no. 4, pp. 600–612, 2004.

[14] C. S. Xydeas and P. V. V., “Objective image fusion performance measure,” Military Technical Courier, vol. 36, no. 4, pp. 308–309, 2000.