VIFB: A Visible and Infrared Image Fusion Benchmark

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

Abstract

Visible and infrared image fusion is an important area in image processing due to its numerous applications. While much progress has been made in recent years with efforts on developing image fusion algorithms, there is a lack of code library and benchmark which can gauge the state-of-the-art. In this paper, after briefly reviewing recent advances of visible and infrared image fusion, we present a visible and infrared image fusion benchmark (VIFB) which consists of 21 image pairs, a code library of 20 fusion algorithms and 13 evaluation metrics. We also carry out extensive experiments within the benchmark to understand the performance of these algorithms. By analyzing qualitative and quantitative results, we identify effective algorithms for robust image fusion and give some observations on the status and future prospects of this field.

1. Introduction

The target of image fusion is to combine information from different images to generate a single image, which is more informative and can facilitate subsequent processing. Many image fusion algorithms have been proposed, which can be generally divided into pixel-level, feature-level and decision-level approaches based on the level of fusion. Also, image fusion can either be performed in the spatial domain or transform domain. Based on application areas, image fusion technology can be grouped into several types, namely medical image fusion [1, 2], multi-focus image fusion [3, 4, 5], remote sensing image fusion [6], multi-exposure image fusion [7, 8], visible and infrared image fusion [9, 10]. Among these types, the visible and infrared image fusion is one of the most frequently used ones. This is because that the visible and infrared image fusion can be applied in many applications, for instance object tracking [11, 12, 13, 14, 15], object detection [16, 17, 18], and biometric recognition [19, 20]. Figure 1 shows an example of visible and infrared image fusion.

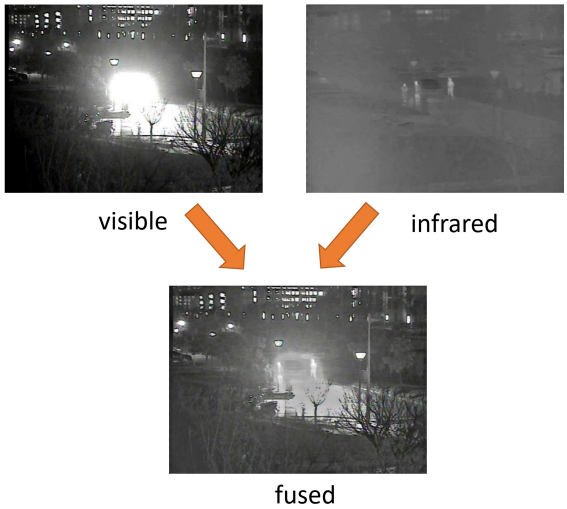

Figure 1. The benefit of visible and infrared image fusion. The people around the car are invisible in visible image due to car light. Although they can be seen in infrared image, the infrared image lacks detail information about the scene. After fusion, the fused image contains enough details and the people are also visible.

However, current research on visible and infrared image fusion is suffering from several problems, which hinder the development of this field severely. First, there is not a well-recognized visible and infrared image fusion dataset which can be used to compare performance under the same standard. Therefore, it is quite common that different images are utilized in experiments in the literature, which makes it difficult to compare the performance of various algorithms. Second, it is crucial to evaluate the performance of state-of-the-art fusion algorithms to demonstrate their strength and weakness and to help identify future research directions in this field. However, although many evaluation metrics have been proposed, none of them is better than all other metrics. As a result, researchers normally just choose several metrics which support their methods in the image fusion literature. This further makes it difficult to objectively compare performances. Third, although the source codes of some image fusion algorithms have been made

Table 1. Details of some existing visible and infrared image fusion datasets and the proposed dataset.

Name	Image/Video pairs	Image type	Resolution	Year	Results	Code library
OSU Color-Thermal Database	6 video pairs	RGB, Infrared	320× 240	2005	No	No
TNO	63 image pairs	multispectral	Various	2014	No	No
VLIRVDIF	24 video pairs	RGB, Infrared	720×480	2019	No	No
VIFB	21 image pairs	RGB, Infrared	Various	2020	Yes	Yes

Figure 2. The infrared and visible test set in VIFB. The dataset includes 21 pairs of infrared and visible images. The first, third, and fifth row contains RGB images, while the second, fourth, and sixth row presents the corresponding infrared images.

publicly available, for example the the DLF [21] and CNN [22], the interface and usage of most algorithms are different and thus it is inconvenient as well as time-consuming for researchers to perform large scale performance evaluation.

To solve these issues, in this work we build a visible and infrared image fusion benchmark (VIFB) that includes 21 pairs of visible and infrared images, 20 publicly available fusion algorithms and 13 evaluation metrics to facilitate the performance evaluation task [1].

The main contributions of this paper lie in the following aspects:

- **Dataset**. We created a test set containing 21 pairs of visible and infrared images. These image pairs are collected from the Internet and several tracking datasets thus covering a wide range of environments and working conditions, such as indoor, outdoor, low illumination, and over-exposure. Therefore, the dataset is able to test the generalization ability of image fusion algorithms.

- **Code library**. We collected 20 recent image fusion algorithms and integrated them into a code library, which can be easily utilized to run algorithms and compare performance. Most of these algorithms are published in recent 5 years. An interface is designed to integrate other image fusion algorithms into VIFB easily.

- **Comprehensive performance evaluation**. We implemented 13 evaluation metrics in VIFB to comprehensively compare fusion performance. We have run the collected 20 algorithms on the proposed dataset and performed comprehensive comparison of those algorithms. All the results are made available for the interested readers to use.

[1] https://github.com/xingchenzhang/Visible-infrared-image-fusion-benchmark

Table 2. Visible and infrared image fusion algorithms that have been integrated in VIFB.

Method	Year	Journal/Conference	Category
ADF [23]	2016	IEEE Sensors Journal	Multi-scale
CBF [24]	2015	Signal, image and video processing	Multi-scale
CNN [22]	2018	International Journal of Wavelets, Multiresolution and Information Processing	DL-based
DLF [21]	2018	International Conference on Pattern Recognition	DL-based
FPDE [25]	2017	International Conference on Information Fusion	Subspace-based
GFCE [26]	2016	Applied Optics	Multi-scale
GFF [27]	2013	IEEE Transactions on Image Processing	Multi-scale
GTF [9]	2016	Information Fusion	Other
HMSD_GF[26]	2016	Applied Optics	Multi-scale
Hybrid_MSD [28]	2016	Information Fusion	Multi-scale
IFEVIP [29]	2017	Infrared Physics & Technology	Other
LatLRR [30]	2018	arXiv	Saliency-based
MGFF [31]	2019	Circuits, Systems, and Signal Processing	Multi-scale
MST_SR [32]	2015	Information Fusion	Hybrid
MSVD [33]	2011	Defense Science Journal	Multi-scale
NSCT_SR [32]	2015	Information Fusion	Hybrid
ResNet [34]	2019	Infrared Physics & Technology	DL-based
RP_SR [32]	2015	Information Fusion	Hybrid
TIF [35]	2016	Infrared Physics & Technology	Saliency-based
VSMWLS [36]	2017	Infrared Physics & Technology	Hybrid

2. Related Work

In this section, we briefly review recent visible and infrared image fusion algorithms. In addition, we summarize existing visible and infrared image datasets.

2.1. Visible-infrared fusion methods

A lot of visible and infrared image fusion methods have been proposed. Before deep learning is introduced to the image fusion community, main image fusion methods can be generally grouped into several categories, namely multi-scale transform-, sparse representation-, subspace-, and saliency-based methods, hybrid models, and other methods according to their corresponding theories [37].

In the past few years, a number of image fusion methods based on deep learning have emerged [38, 39, 40, 37]. Deep learning can help to solve several important problems in image fusion. For example, deep learning can provide better features compared to handcrafted ones. Besides, deep learning can learn adaptive weights in image fusion, which is crucial in many fusion rules. Regarding methods, convolutional neural network (CNN) [41, 4, 5, 2, 8], generative adversarial networks (GAN) [42], Siamese networks [22], autoencoder [43] have been explored to conduct image fusion. Apart from image fusion methods, the image quality assessment, which is critical in image fusion performance evaluation, has also benefited from deep learning [44]. It is foreseeable that image fusion technology will develop in the direction of machine learning, and an increasing number of research results will appear in the coming years.

2.2. Existing dataset

Although the research on image fusion has begun for many years, there is still not a well-recognized and commonly used dataset in the community of visible and infrared image fusion. This differs from the visual tracking community where several well-known benchmarks have been proposed and widely utilized, such as OTB [45, 46] and VOT [47]. Therefore, it is common that different image pairs are utilized in visible and infrared image fusion literature, which makes the objective comparison difficult.

At the moment, there are several existing visible and infrared image fusion datasets, including OSU Color-Thermal Database [48][2], TNO Image fusion dataset[3], and VLIRVDIF [49][4]. The main information about these datasets are summarized in Table 1. Actually, apart from OSU, the number of image pairs in TNO and VLIRVDIF is not small. However, the lack of code library, evaluation metrics as well as results on these datasets make it difficult to gauge the state-of-the-art based on them.

3. Visible and Infrared Image Fusion Benchmark

3.1. Dataset

The dataset in VIFB, which is a test set, includes 21 pairs of visible and infrared images. The images are collected by the authors from the Internet[5] and fusion tracking dataset [62, 48, 13]. These images cover a wide range of environments and working conditions, such as indoor, outdoor, low illumination, and over-exposure. Each pair of visible and infrared image has been registered to make sure that the image fusion can be successfully performed. There are various im-

[2]http://vcipl-okstate.org/pbvs/bench/

[3]https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029

[4]http://www02.smt.ufrj.br/ fusion/

[5]https://www.ino.ca/en/solutions/video-analytics-dataset/

Table 3. Evaluation metrics implemented in VIFB. '+' means that a large value indicates a good performance while '-' means that a small value indicates a good performance.

Category	Name	Meaning	+/-	Category	Name	Meaning	+/-
Information theory-based	CE [50]	Cross entropy	-	Image feature-based	AG [51]	Average gradient	+
	EN [52]	Entropy	+		EI [53]	Edge intensity	+
	MI [54]	Mutual information	+		SD [55]	Standard deviation	+
	PSNR [56]	Peak signal-to-noise ration	+		SF [57]	Spatial frequency	+
					$Q^{AB/F}$ [58]	Gradient-based fusion performance	+
Structural similarity-based	SSIM [59]	Structural similarity index measure	+	Human perception inspired	Q_{CB} [60]	Chen-Blum metric	+
	RMSE [56]	Root mean squared error	-		Q_{CV} [61]	Chen-Varshney metric	-

Figure 3. Qualitative comparison of 20 methods on the *fight* image pair shown in Fig. 2.

age resolution in the dataset, such as 320×240, 630×460, 512×184, and 452×332. Some examples of images in the dataset are given in Fig. 2.

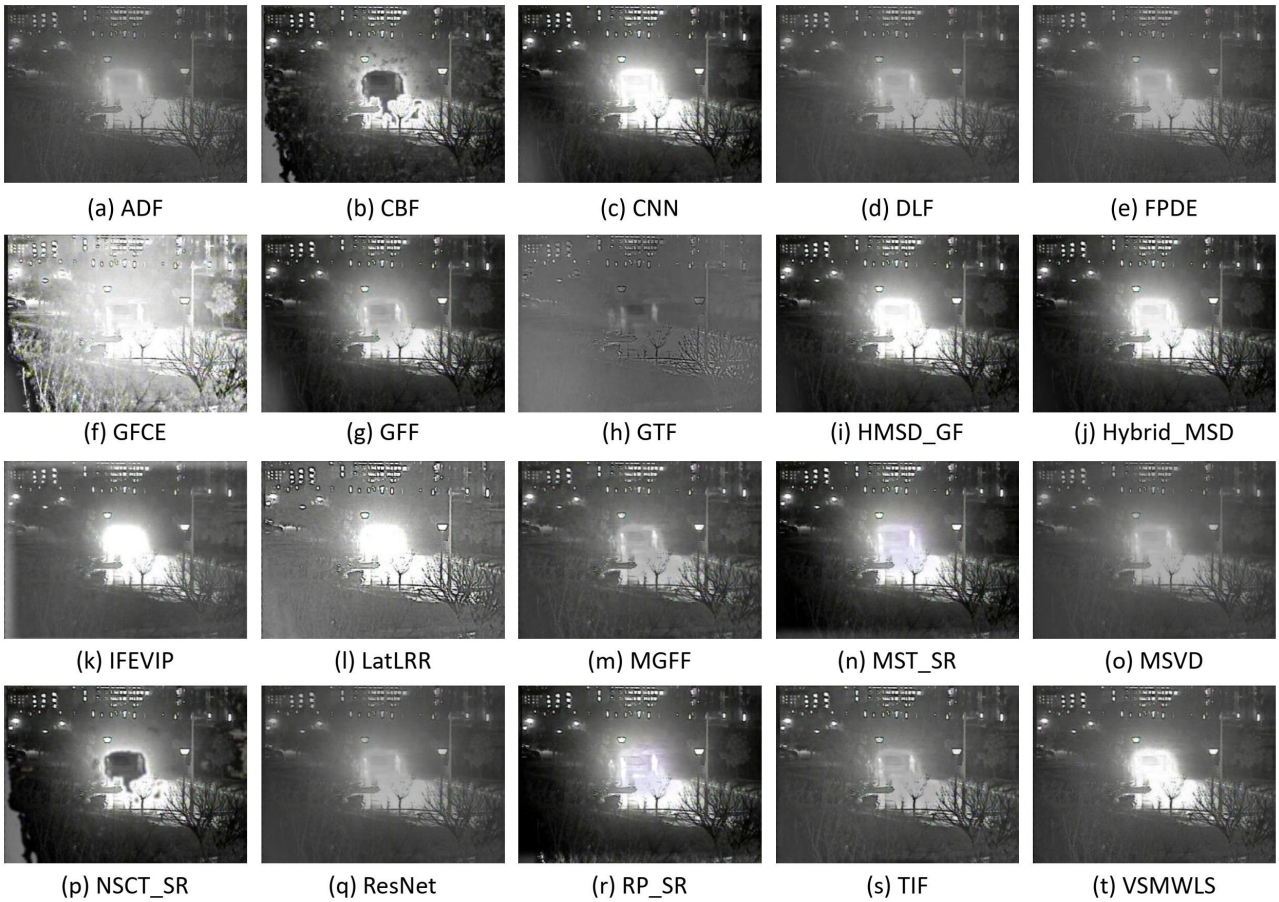

Figure 4. Qualitative comparison of 20 methods on the *manlight* image pair shown in Fig. 1 and Fig. 2.

3.2. Baseline algorithms

In recent years, a lot of algorithms have been proposed to perform visible and infrared image fusion. However, only a part of papers provide the source code. Besides, these codes have different input and output interfaces, and they may require different running environment. These factors hinder the usage of these codes to produce results and to perform large-scale performance comparison.

In VIFB benchmark, we integrated 20 recently published visible-infrared image fusion algorithms including MSVD [33], GFF [27], MST_SR [32], RP_SR [32], NSCT_SR [32], CBF [24], ADF [23], GFCE [26], HMSD_GF [26], Hybrid-MSD [28], TIF [35], GTF [9], FPDE [25], IFEVIP [29], VSM_WLS [36], DLF [21], LatLRR [30], CNN [22], MGFF [31], ResNet [34]. Table 2 lists more details about these algorithms. Note that many algorithms were originally designed to fuse grayscale images. We modified them to fuse color images by fusing every channel of the RGB image with corresponding infrared image.

These algorithms cover almost every kind of visible-infrared fusion algorithms, and most algorithms are proposed in the last five years, which can represent the development of the visible-infrared fusion field to some extent.

To integrate algorithms into VIFB and for the convenience of users, we designed an interface. By using this interface, other visible-infrared fusion algorithms or their fusion results can be integrated to VIFB to compare their results with those integrated algorithms.

3.3. Evaluation metrics

Numerous evaluation metrics for visible-infrared image fusion have been proposed. As introduced in [63], image fusion metrics can be classified into four types, namely information theory-based, image feature-based, image structural similarity-based, and human perception-based metrics. However, none of the proposed metrics is better than all others. To have comprehensive and objective performance comparison, we implemented 13 evaluation metrics in VIFB. All evaluation metrics that have been implemented in VIFB and their corresponding categories are listed in Table 3. As can be seen, the implemented metrics in VIFB cover all four categories. It is convenient to compute all these metrics for each method in VIFB, thus making it easy

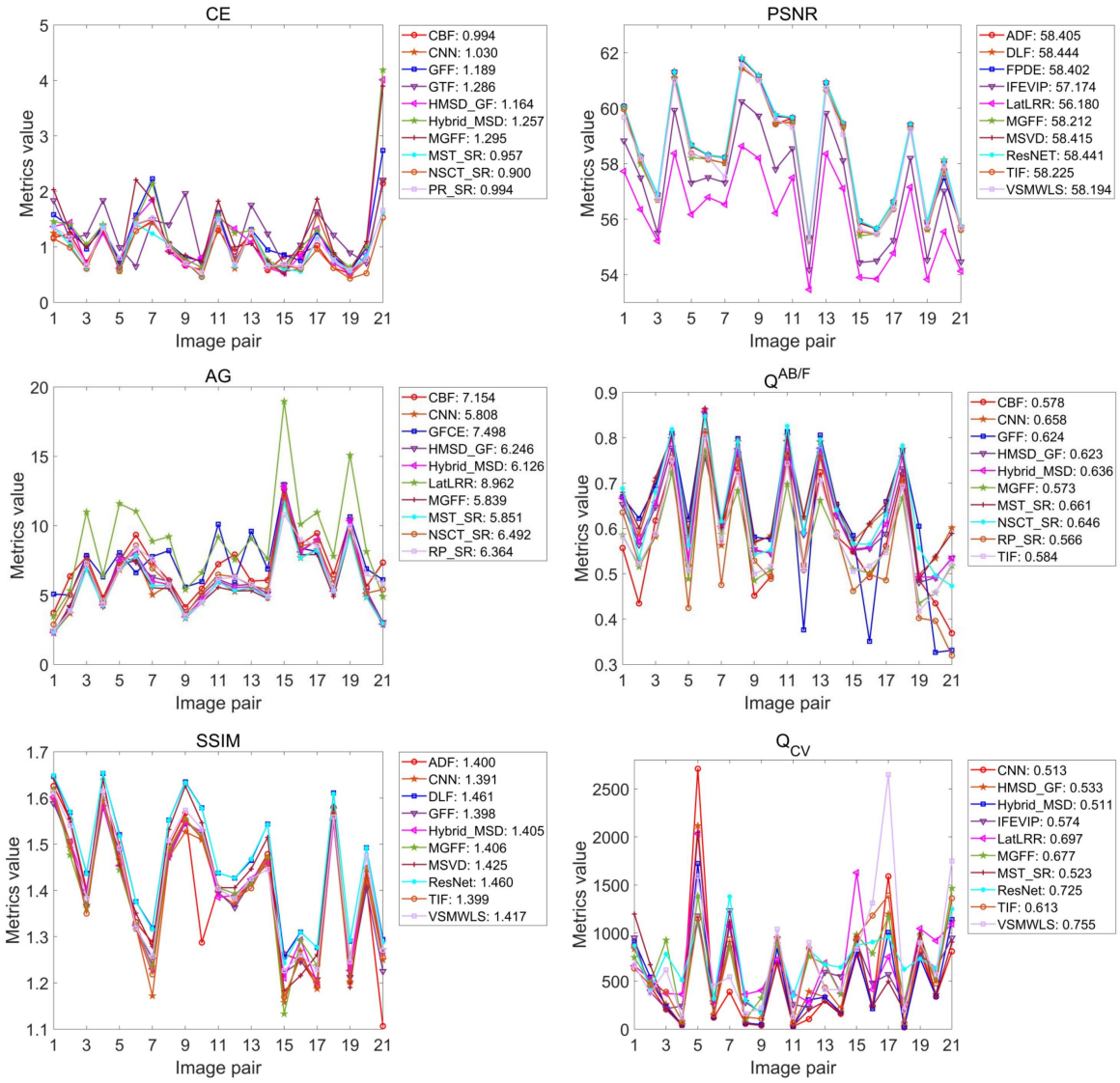

Figure 5. Quantitative comparisons of six metrics of the selected 10 methods on 21 image pairs shown in Fig. 2. The best 10 methods in terms of each evaluation metric are shown. The values in the legend indicate the average value on 21 image pairs for each method. From 1 to 21 in the horizontal axis: *carLight, carShadow, carWhite, elecbike, fight, kettle, labMan, man, manCall, manCar, manlight, manWalking, manwithbag, nightCar, peopleshadow, running, snow, tricycle, walking, walking2, walkingnight*.

to compare performances among methods. Due to the page limits, we leave the detailed introduction to these metrics in the supplementary material. More information about evaluation metrics can be founded in [63, 37].

4. Experiments

This section presents experimental results on the VIFB dataset. Section 4.1 and Section 4.2 presents qualitative and quantitative performance comparison, respectively. Section 4.3 compares the runtime of each algorithm. All experiments were performed using a computer equipped with an NVIDIA RTX2070 GPU and i7-9750H CPU. Default parameters reported by the corresponding authors of each algorithm were employed. Regarding deep learning-based algorithms, the pretrained models provided by their authors were used in this work and we did not retrain those models. Note that due to the page limits, we just present a part of results here. More fusion results will be provided in the supplementary materials.

4.1. Qualitative performance comparison

Qualitative evaluation methods are important in fusion quality assessment and they assess the quality of fused images on the basis of the human visual system. Figure 3

Table 4. Average evaluation metric values of all methods on 21 image pairs. The best three values in each metric are denoted in red, green and blue, respectively. The three numbers after the name of each method denote the number of best value, second best value and third best value, respectively. Best viewed in color.

Method	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	Q_{CB}	Q_{CV}	RMSE	SF	SSIM	SD
ADF (0,0,0)	4.582	1.464	46.529	6.788	1.921	58.405	0.520	0.474	777.8	0.1043	14.132	1.400	35.185
CBF (0,0,3)	7.154	0.994	74.590	7.324	2.161	57.595	0.578	0.526	1575.1	0.1257	20.380	1.171	48.544
CNN (1,2,2)	5.808	1.030	60.241	7.320	2.653	57.932	0.658	0.622	512.6	0.1178	18.813	1.391	60.075
DLF (3,0,0)	3.825	1.413	38.569	6.724	2.030	58.444	0.434	0.445	759.8	0.1035	12.491	1.461	34.717
FPDE (0,0,0)	4.538	1.366	46.022	6.766	1.924	58.402	0.484	0.460	780.1	0.1045	13.468	1.387	34.931
GFCE (0,3,0)	7.498	1.931	77.466	7.266	1.844	55.939	0.471	0.535	898.9	0.1728	22.463	1.134	51.563
GFF (0,0,0)	5.326	1.189	55.198	7.210	2.638	58.100	0.624	0.619	881.6	0.1117	17.272	1.398	50.059
GTF (0,0,0)	4.303	1.286	43.664	6.508	1.991	57.861	0.439	0.414	2138.4	0.1177	14.743	1.371	35.130
HMSD_GF (0,1,0)	6.246	1.164	65.034	7.274	2.472	57.940	0.623	0.604	533.0	19.904	19.904	1.394	57.617
Hybrid_MSD (1,1,0)	6.126	1.257	63.491	7.304	2.619	58.173	0.636	0.623	510.9	0.1102	19.659	1.405	54.922
IFEVIP (0,0,0)	4.984	1.339	51.782	6.936	2.248	57.174	0.486	0.462	573.8	0.1384	15.846	1.391	48.491
LatLRR (3,0,1)	8.962	1.684	92.813	6.909	1.653	56.180	0.438	0.497	697.3	0.1686	29.537	1.184	57.134
MGFF (0,0,0)	5.839	1.295	60.607	7.114	1.768	58.212	0.573	0.542	676.9	0.1092	17.916	1.406	44.290
MST_SR (2,2,2)	5.851	0.957	60.781	7.339	2.809	57.951	0.661	0.645	522.7	0.1165	18.807	1.390	57.314
MSVD (0,0,3)	3.545	1.462	36.202	6.705	1.955	58.415	0.332	0.426	809.0	0.1042	12.525	1.425	34.372
NSCT_SR (3,1,0)	6.492	0.900	67.956	7.396	2.988	57.435	0.646	0.617	1447.3	0.1314	19.389	1.277	52.475
ResNet (1,2,0)	3.674	1.364	37.255	6.735	1.988	58.441	0.407	0.445	724.8	0.1035	11.736	1.460	34.940
RP_SR (0,1,1)	6.364	0.994	65.219	7.353	2.336	57.777	0.566	0.606	888.8	0.1217	21.171	1.332	55.808
TIF (0,0,0)	5.558	1.371	57.839	7.075	1.767	58.225	0.584	0.545	613.0	0.1087	17.739	1.399	42.643
VSMWLS (0,0,0)	5.612	1.409	57.252	7.028	2.035	58.194	0.554	0.497	754.7	0.1092	17.662	1.417	46.253

presents the qualitative performance comparison of 20 fusion methods on the *fight* image pair. In this image pair, several people are in the shadow of a car thus can not be seen clearly in the visible image while can be seen in infrared image. As can be seen, in almost all fused images these people can be seen. However, the fused images obtained by some algorithms have introduced artifacts information. These include CBF, IFEVIP, MST_SR, NSCT_SR, and RP_SR. Besides, the fused images produced by ADF, CNN, GTF, LatLRR and MSVD do not preserve detail information contained in the visible image well. Figure 3 indicates that the fused images obtained by Hybrid_MSD, MGFF, TIF and VSMWLS are more natural for human sensitivity and preserve more details.

Figure 4 shows the qualitative comparison of 20 methods on *manlight* image pair. In this case, the people around the car are invisible in visible images due to over-exposure. It can be seen that in many fused images, the people around the car are still invisible or not clear, such as those produced by CNN, GFCE, HMSD_GF, Hybrid_MSD, IFEVIP, LatLRR, and VSMWLS. Some other fused images have more artifacts which are not presented in original images, such as those obtained by CBF, GFCE, and NSCT_SR. Although the fused images produced by MST_SR and RP_SR preserve the details in visible image well and the people around the car can be seen clearly, some light purple are introduced in the fused images (near the image center) which are not presented in source images. The results indicate that GFF and MGFF give better subjective fusion performance for the *manlight* case.

4.2. Quantitative performance comparison

Table 4 presents the average value of 13 evaluation metrics for all methods on 21 image pairs. As can be seen, the NSCT_SR obtains the best overall quantitative performance by having 3 best values and 1 second best value. The LatLRR method obtains the second best overall performance by having 3 best values and 1 third best value. DLF ranks the third place in terms of overall performance by having 3 best values. However, this table indicates clearly that there is not a dominant fusion method that can beat other methods in all or most evaluation metrics. Besides, from the table one can see that the deep learning-based methods show slightly worse performance than conventional fusion algorithms, although each deep learning-based method performs well in some evaluation metrics. This is very different from the field of object tracking and detection which is almost dominated by deep learning-based approaches.

From Table 4 one can also see that the top three algorithms show very different performance in different kinds of metrics. Specifically, the NSCT_SR algorithm obtains the best value in CE, EN and MI, which are all information theory-based evaluation metrics. The LatLRR algorithm shows the best performance in AG, EI and SF, which are all image feature-based metrics. The DLF method performs well in RMSE, SSIM and PSNR. Both RMSE and SSIM are structural similarity-based metrics. The possible reason is that the authors of these algorithms pay more attention to a specific kind of information when designing these algorithms. This phenomenon further shows that an image fusion algorithm should be evaluated using various kinds of metrics for a comprehensive comparison, which

further indicates the benefits of this study.

Note that although the NSCT_SR algorithm obtains the best overall quantitative performance, its qualitative performance is not very good. As can be seen from Fig. 3 and Fig. 4, it introduces artifacts in the fused images. Similarly, the LatLRR also shows good quantitative performance but the qualitative performance is relatively poor. Specifically, in the *fight* case the LatLRR algorithm loses some details of the visible image while in the *manlight* case it fails to show the target which is invisible in visible images due to over-exposure. Actually, NSCT_SR and LatLRR do not perform very well in Q_{CB} and Q_{CV}, which are human perception inspired metrics used to measure the visual performance of the fused image. The different performance between qualitative and quantitative evaluation clearly shows that both qualitative and quantitative comparison are crucial in image fusion quality evaluation.

To further show quantitative comparison of fusion performances of various methods, the values of six metrics of the 10 selected methods on 21 image pairs are presented in Figure 5.

4.3. Runtime comparison

The runtime of algorithms integrated in VIFB is listed in Table 5. As can be seen, the runtime of image fusion methods varies significantly from one to another. This is also true even for methods in the same category. For instance, both CBF and GFF are multi-scale methods, but the runtime of CBF is more than 50 times that of GFF. Besides, multi-scale methods are generally fast and deep learning-based algorithms are slower than others even with the help of GPU. The fastest deep learning-based method, i.e. ResNet, takes 4.80 seconds to fuse one image pair. It should be mentioned that all three deep learning-based algorithms in VIFB do not update the model online, but use pretrained model instead.

One important application area of visible and infrared image fusion is the RGB-infrared fusion tracking [11, 12, 64], where the tracking speed is vital for practical applications. As pointed out in [11], if an image fusion algorithm is very time-consuming, like LatLRR [30] and NSCT_SR [32], then it will not be feasible to develop a real-time fusion tracker based on this image fusion algorithm. Actually, most image fusion algorithms listed in Table 5 are computationally expensive in terms of tracking.

Table 5. Runtime of algorithms in VIFB (seconds per image pair)

Method	Average runtime	Category
ADF [23]	1.00	Multi-scale
CBF [24]	22.97	Multi-scale
CNN [22]	31.76	DL-based
DLF [21]	18.62	DL-based
FPDE [25]	2.72	Subspace-based
GFCE [26]	2.13	Multi-scale
GFF [27]	0.41	Multi-scale
GTF [9]	6.27	Other
HMSD_GF[26]	2.76	Multi-scale
Hybrid_MSD [28]	9.04	Multi-scale
IFEVIP [29]	0.17	Other
LatLRR [30]	271.04	Saliency-based
MGFF [31]	1.08	Multi-scale
MST_SR [32]	0.76	Hybrid
MSVD [33]	1.06	Multi-scale
NSCT_SR [32]	94.65	Hybrid
ResNet [34]	4.80	DL-based
RP_SR [32]	0.86	Hybrid
TIF [35]	0.13	Saliency-based
VSMWLS [36]	3.51	Hybrid

5. Concluding Remarks

In this paper, we present a visible and infrared image fusion benchmark (VIFB), which includes a test set of 21 image pairs, a code library consists of 20 algorithms, 13 evaluation metrics and all results. To the best of our knowledge, this is the first visible and infrared image fusion benchmark to date. This benchmark facilitates better understanding of the state-of-the-art image fusion approaches, and can provide a platform for gauging new methods.

We carry out extensive experiments using VIFB to evaluate the performance of all integrated fusion algorithms. We have several observations based on our experimental results. First, unlike some other fields in computer vision where deep learning is almost dominant, such as object tracking and detection, the performances of deep learning-based image fusion algorithms do not show superiority over non-learning algorithms at the moment. However, due to its strong representation ability, we believe that the deep learning-based image fusion approach will be an important research direction in future. Second, image fusion algorithms may have different performances in different kinds of evaluation metrics, thus it is necessary to utilize various kinds of metrics to comprehensively evaluate an image fusion algorithm. Besides, both qualitative and quantitative evaluation are crucial. Finally, the computational efficiency of visible and infrared image fusion algorithms still need to be improved in order to be applied in real-time applications, such as tracking and detection.

We will continue extending the dataset and code library of VIFB. We will also implement more evaluation metrics in VIFB. We hope that VIFB can serve as a good starting point for researchers who are interested in visible and infrared image fusion.

Acknowledgments. This work was sponsored in part by the National Natural Science Foundation of China under Grant 61973212, in part by the Shanghai Science and Technology Committee Research Project under Grant 17DZ1204304.

References

[1] A. P. James and B. V. Dasarathy, "Medical image fusion: A survey of the state of the art," Information Fusion, vol. 19, pp. 4–19, 2014.

[2] K. Xia, H. Yin, and J. Wang, "A novel improved deep convolutional neural network model for medical image fusion," Cluster Computing, pp. 1–13, 2018.

[3] Z. Wang, Y. Ma, and J. Gu, "Multi-focus image fusion using pcnn," Pattern Recognition, vol. 43, no. 6, pp. 2003–2016, 2010.

[4] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," Information Fusion, vol. 36, pp. 191–207, 2017.

[5] X. Yan, S. Z. Gilani, H. Qin, and A. Mian, "Unsupervised deep multi-focus image fusion," arXiv preprint arXiv:1806.07272, 2018.

[6] H. Ghassemian, "A review of remote sensing image fusion methods," Information Fusion, vol. 32, pp. 75–89, 2016.

[7] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3345–3356, 2015.

[8] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, "Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017, pp. 4724–4732.

[9] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," Information Fusion, vol. 31, pp. 100–109, 2016.

[10] D. P. Bavirisetti, G. Xiao, J. Zhao, X. Zhang, and P. Wang, "A new image and video fusion method based on cross bilateral filter," in 2018 21st International Conference on Information Fusion (FUSION). IEEE, 2018, pp. 1–8.

[11] X. Zhang, G. Xiao, P. Ye, D. Qiao, J. Zhao, and S. Peng, "Object fusion tracking based on visible and infrared images using fully convolutional siamese networks," in Proceedings of the 22nd International Conference on Information Fusion. IEEE, 2019.

[12] X. Zhang, P. Ye, S. Peng, J. Liu, K. Gong, and G. Xiao, "SiamFT: An RGB-infrared Fusion Tracking Method via Fully Convolutional Siamese Networks," IEEE Access, vol. 7, pp. 122 122–122 133, 2019.

[13] C. Li, X. Liang, Y. Lu, N. Zhao, and J. Tang, "Rgb-t object tracking: benchmark and baseline," Pattern Recognition, p. 106977, 2019.

[14] N. Xu, G. Xiao, X. Zhang, and D. P. Bavirisetti, "Relative object tracking algorithm based on convolutional neural network for visible and infrared video sequences," in Proceedings of the 4th International Conference on Virtual Reality. ACM, 2018, pp. 44–49.

[15] N. Xu, G. Xiao, F. He, X. Zhang, and D. P. Bavirisetti, "Object tracking via deep multi-view compressive model for visible and infrared sequences," in Proceedings of the 21st International Conference on Information Fusion (FUSION). IEEE, 2018, pp. 941–948.

[16] H. Torresan, B. Turgeon, C. Ibarra-Castanedo, P. Hebert, and X. P. Maldague, "Advanced surveillance systems: combining video and thermal imagery for pedestrian detection," in Thermosense XXVI, vol. 5405. International Society for Optics and Photonics, 2004, pp. 506–516.

[17] R. Lahmyed, M. El Ansari, and A. Ellahyani, "A new thermal infrared and visible spectrum images-based pedestrian detection system," Multimedia Tools and Applications, vol. 78, no. 12, pp. 15 861–15 885, 2019.

[18] Y. Yan, J. Ren, H. Zhao, G. Sun, Z. Wang, J. Zheng, S. Marshall, and J. Soraghan, "Cognitive fusion of thermal and visible imagery for effective detection and tracking of pedestrians in videos," Cognitive Computation, vol. 10, no. 1, pp. 94–104, 2018.

[19] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognition - A review," Computer Vision and Image Understanding, vol. 97, no. 1, pp. 103–135, 2005.

[20] S. M. Z. S. Z. Ariffin, N. Jamil, P. N. M. A. Rahman, S. Mohd, Z. Syed, Z. Ariffin, N. Jamil, and U. Mara, "Can thermal and visible image fusion improves ear recognition?" in Proceedings of the 8th International Conference on Information Technology, 2017, pp. 780–784.

[21] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," 24th International Conference on Pattern Recognition, 2018.

[22] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," International Journal of Wavelets, Multiresolution and Information Processing, vol. 16, no. 03, p. 1850018, 2018.

[23] D. P. Bavirisetti and R. Dhuli, "Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform," IEEE Sensors Journal, vol. 16, no. 1, pp. 203–209, 2016.

[24] B. K. Shreyamsha Kumar, "Image fusion based on pixel significance using cross bilateral filter," Signal, Image and Video Processing, vol. 9, no. 5, pp. 1193–1204, Jul 2015.

[25] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in 2017 20th International Conference on Information Fusion (Fusion). IEEE, 2017, pp. 1–9.

[26] Z. Zhou, M. Dong, X. Xie, and Z. Gao, "Fusion of infrared and visible images for night-vision context enhancement," Applied optics, vol. 55, no. 23, pp. 6480–6490, 2016.

[27] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," IEEE Transactions on Image processing, vol. 22, no. 7, pp. 2864–2875, 2013.

[28] Z. Zhou, B. Wang, S. Li, and M. Dong, "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters," Information Fusion, vol. 30, pp. 15–26, 2016.

[29] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, "Infrared and visual image fusion through infrared feature extraction and visual information preservation," Infrared Physics & Technology, vol. 83, pp. 227 – 237, 2017.

[30] H. Li and X. Wu, "Infrared and visible image fusion using latent low-rank representation," arXiv preprint arXiv:1804.08992, 2018.

[31] D. P. Bavirisetti, G. Xiao, J. Zhao, R. Dhuli, and G. Liu, "Multi-scale guided image and video fusion: A fast and efficient approach," Circuits, Systems, and Signal Processing, vol. 38, no. 12, pp. 5576–5605, Dec 2019.

[32] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," Information Fusion, vol. 24, pp. 147–164, 2015.

[33] V. Naidu, "Image fusion technique using multi-resolution singular value decomposition," Defence Science Journal, vol. 61, no. 5, pp. 479–484, 2011.

[34] H. Li, X.-J. Wu, and T. S. Durrani, "Infrared and visible image fusion with resnet and zero-phase component analysis," Infrared Physics & Technology, vol. 102, p. 103039, 2019.

[35] D. P. Bavirisetti and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection," Infrared Physics & Technology, vol. 76, pp. 52–64, 2016.

[36] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," Infrared Physics & Technology, vol. 82, pp. 8–17, 2017.

[37] 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.

[38] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, J. Hai, and K. He, "A survey of infrared and visual image fusion methods," Infrared Physics & Technology, vol. 85, pp. 478–501, 2017.

[39] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," Information Fusion, vol. 33, pp. 100–112, 2017.

[40] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," Information Fusion, vol. 42, pp. 158–173, 2018.

[41] H. Hermessi, O. Mourali, and E. Zagrouba, "Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain," Neural Computing and Applications, pp. 1–17, 2018.

[42] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," Information Fusion, vol. 48, no. June 2018, pp. 11–26, 2019.

[43] H. Li and X. Wu, "Densefuse: A fusion approach to infrared and visible images," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2614–2623, 2018.

[44] Q. Yan, D. Gong, and Y. Zhang, "Two-stream convolutional networks for blind image quality assessment," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2200–2211, 2018.

[45] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2411–2418.

[46] ——, "Object tracking benchmark," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1834–1848, 2015.

[47] M. Kristan, J. Matas, A. Leonardis, T. Vojir, R. Pflugfelder, G. Fernandez, G. Nebehay, F. Porikli, and L. Čehovin, "A novel performance evaluation methodology for single-target trackers," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 11, pp. 2137–2155, Nov 2016.

[48] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," Computer vision and image understanding, vol. 106, no. 2-3, pp. 162–182, 2007.

[49] A. Ellmauthaler, C. L. Pagliari, E. A. da Silva, J. N. Gois, and S. R. Neves, "A visible-light and infrared video database for performance evaluation of video/image fusion methods," Multidimensional Systems and Signal Processing, vol. 30, no. 1, pp. 119–143, 2019.

[50] 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.

[51] 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.

[52] 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.

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

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

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

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

[57] 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.

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

[59] 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.

[60] 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.

[61] 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.

[62] C. O'Conaire, N. E. O'Connor, E. Cooke, and A. F. Smeaton, "Comparison of fusion methods for thermo-visual surveillance tracking," in 2006 9th International Conference on Information Fusion. IEEE, 2006, pp. 1–7.

[63] 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.

[64] X. Zhang, P. Ye, S. Peng, J. Liu, and G. Xiao, "DSiamMFT: An RGB-T fusion tracking method via dynamic Siamese networks using multi-layer feature fusion," Signal Processing: Image Communication, p. 115756, 2020.