2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) | 979-8-3503-8592-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONECCT62155.2024.10677151

A Colour-Focused Visible and Infrared Image Fusion Framework for Aiding Human Perception

Nithin Eswarappa[1*], Shefali Waldekar[2†], Jeevan K. M.[1‡], Bikram Kumar Vivek[1§] and Koshy George[1¶]
[1]Department of EECE, Gandhi Institute of Technology and Management (GITAM), Bangalore, India
[2]Department of ECE, Nirma University, Ahmedabad, India
Email: *neswarap@gitam.in, †shefali.waldekar@nirmauni.ac.in, ‡jmani@gitam.edu,
§bvivek@gitam.in, ¶kgeorge@{gitam.edu,ieee.org}

***Abstract*—Image fusion methods are required when images from the same or different sensors are available. In several applications, photographs from a colour camera are often fused with thermal infrared (medium-wave or long-wave) images. This paper focuses on those applications (typically security and surveillance) wherein the fused images are used by a human operator who will, in turn, make critical decisions. This paper presents two frameworks for fusing colour and infrared images, one in the red-green-blue (RGB) colour space and the other in lightness-chroma-hue (LCH) colour space. We compare qualitatively and quantitatively the performance of five existing fusion methods that work best in this scenario. The study is carried out across two datasets. Quantitatively, four existing metrics are used to analyse the performance. In addition, this paper proposes a performance metric that appeals to human perception and measures how well colour, edge, and contrast information are transferred to the fused image.**

***Index Terms*—Image Fusion, Visible Infrared Image Fusion, Fusion Performance Metrics, Quantitative Metrics**

I. Introduction

Human vision forms a significant component in environmental perception. It has adapted to respond to the visible spectrum of light. However, an environment's perception is not limited to the visible spectrum. Materials are distinguishable based on their transmittance, absorptance, and reflectance of various electromagnetic spectrum frequencies. An optical sensor responds to a window of wavelengths due to limiting factors arising from sensor design to choice of sensor material. Visible images offer rich colour and texture information, but the quality depends on illumination and environmental conditions such as smoke, fog, rain, and other particulate matter. Thermal infrared images are independent of lighting and offer better visibility in challenging environmental conditions. Optical sensors — ultraviolet, visible, infrared, and Synthetic Aperture Radar (SAR) — are band-limited in their response, leading to a narrow perception of available information. In remote sensing, one sensor offers high spectral details, whereas the other offers images with better spatial resolution. In medical imaging, Computed Tomography (CT) images give structural details, whereas Magnetic Resonance Imaging (MRI) have functional information. Multiple applications, such as remote sensing, surveillance, multi-focus image fusion, medical imaging, multi-exposure image fusion (HDR imaging), astronomical imaging, robotics, industrial inspection, multimedia, and bio-metrics, use image fusion methods to combine the complementary properties of sensors (e.g., visible and infrared image fusion (VIIF) methods) or images (e.g., multi-focus image fusion (MFIF) methods) to yield a more informative fused image.

Image fusion involves generating a single fused image (grey-scale/colour) using either source images or combining the source images based on a fusion rule to maximise available perceivable information and simultaneously minimise loss of information from source images. Based on the abstraction level of an image used for fusion, these methods have been classified into pixel-level, feature-level, and decision-level image fusion [1]. Multimodal image fusion techniques, in general, and visible and infrared image fusion methods, in particular, are well-studied areas in image processing. Image fusion is carried out, by transforming the source images based on domains such as multi-scale decomposition, multi-scale geometric analysis, sparse representation, subspace, IHS transformations and using neural networks methods, saliency detection methods, and hybrid methods. In [2], the authors created a benchmark for VIIF. It contains a toolbox with 20 fusion methods, 13 evaluation metrics, along with 21 registered image pairs. They noted that multi-scale transform methods with sparse representations performed the best quantitatively and qualitatively and had faster compute times. In general, multi-scale methods were faster than algorithms based on deep learning, even with the help of a GPU. In remote sensing and visible infrared image fusion applications, fusion methods that used colour space transformation have been explored [3]. IHS transform fusion methods are easy to implement [4] as they sometimes replace the I-channel of the visible image, which has a low spatial resolution, with an image with a high spatial resolution.

In applications such as surveillance, a live data feed of fused images needs to be available. Although much needed, real-time processing and fusion of multimodal (visible and infrared) images, particularly for deep learning-based methods are challenging on embedded platforms with multi-sensor configurations [5]. In [6], it was observed that deep learning-based MFIF approaches did not show advantages over conventional methods. Accordingly, this paper chooses traditional image fusion methods over machine and deep learning-based

methods for comparative analysis. Image fusion methods are evaluated based on performance metrics [7]. These performance metrics have been classified into two categories: applications where reference images are available, and applications where they are not. VIIF methods come under the latter. In such cases, measuring the performance of the fusion method becomes challenging. Usually, the performances of fusion methods are presented in two categories: qualitative analysis (e.g., Wald's protocol) and quantitative analysis. Fusion performance evaluated using quantitative methods [8] do not always corroborate with observations from qualitative analysis [9]. Therefore, we propose a new metric, called the *colour fusion metric*, that aims to capture the details of human vision perception that will correlate with qualitative analysis. This measure accounts for colour information transferred from the visible source image as well as edge and contrast information [10].

This paper presents a comparative study of VIIF methods using two frameworks based on colour spaces. Unlike most existing techniques, the objective is to provide comprehensive information to a human operator in the loop. Such applications typically arise in security and surveillance. Thus, the purpose of image fusion here is to aid a human operator in making effective decisions. Typical metrics to evaluate image fusion performance do not have this perspective. In contrast to this, the novel colour fusion metric mentioned earlier estimates the efficacy of transferring colour, edge, and contrast information. The framework and the metric is tested with two datasets: (a) M^3FD Dataset [11] and (b) LLVIP Dataset [12]. Further details on these datasets are presented in Section III.

The rest of the paper is organised as follows. The image fusion frameworks and the proposed fusion performance metric are described in Section II. The experimental setup, datasets, and results are presented and discussed in Section III, followed by conclusions in Section IV.

II. Methodology

The typical steps in generating a fused image using conventional VIIF methods are as follows: (a) Registration of source images. (b) Preprocessing. (c) Applying a fusion rule at a selected image abstraction (pixel-level, feature-level, or decision-level). (d) Fused image construction.

In the context of this paper, two paths are explored for fusing colour and infrared images. The frameworks for RGB and LCH colour spaces are shown in Fig. 1 and Fig. 2, respectively. In the former case, fusion methods are applied directly and independently on each channel and combined to form the final fused colour image. The latter scenario transforms the RGB image into the LCH colour space. Only the lightness channel of the image is fused with the infrared image. The resulting image is then converted back to the RGB colour space to obtain the final fused image. To our knowledge, there appears to be no related work in the LCH colour space. Compared to the hue-saturation-value (HSV) colour space, the LCH colour space is perceptually more

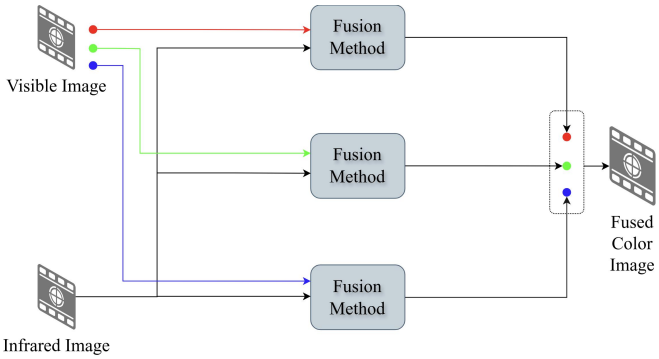


Fig. 1: RGB colour space fusion framework

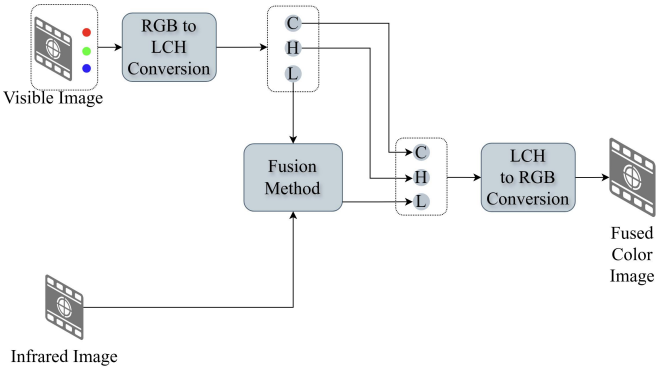


Fig. 2: LCH colour space fusion framework.

uniform. That is, a difference in hue values yields the same visual colour difference.

A. Proposed Colour Fusion Metric (CFM)

The proposed evaluation metric has been formulated to achieve a correlation with a qualitative assessment of image fusion. Three components that appeal to human perception are selected: colour, edge, and contrast. This metric helps in evaluating the fusion method for applications where the fused image is viewed by a person to aid in decision-making. This fusion evaluation metric requires the source images to be registered. It has three components: The first estimates the transfer of hue information from visible to fused image, the second estimates the preservation of edges from both visible and infrared modes along with the strength of the edges, while the third is a global contrast measure [10]. We first convert the visible image from RGB colour space to LCH colour space. The edge, gradient and contrast information is extracted from the lightness channel. Chroma and hue are used to estimate the transfer of colour information.

Let the matrices H_{vis} and H_{fus} represent the hue channels of the visible and fused images, respectively. The first computational step of the proposed metric is the following ratio:

$$\text{Rate}_{\text{hue}} = d_{\text{hue}} \odot (H_{\text{vis}})^{-1} \tag{1}$$

where $\odot$ is the Hadamard product (denoting element-by-element product), and the zero elements of H_{vis} are replaced by unity to avoid division by zero. Moreover, d_{hue} is a matrix

of differences between the hue values of the visible and fused images:

$$d_{\text{hue}} = |H_{\text{vis}} - H_{\text{fus}}| \odot C_{\text{vis}}, \tag{2}$$

where the computations for those pixels with low intensities are not required, as these pixels would be replaced by the corresponding values from the IR image during the fusion process. Moreover, these computations make sense for those pixels which contain significant colour information. In (2), C_{vis} represents the chroma channel of the visible image.

For computing the edge component of the performance metric, we need edge and gradient images from the source and fused images. The gradient of the source image is computed as $G_{\text{src}} = \max\left\{G_{\text{L}_{\text{vis}}}, G_{\text{ir}}\right\}$, where $G_{\text{L}_{\text{vis}}}$ and G_{ir} are respectively the gradients of visible and infrared images obtained using the Roberts gradient operator. The edge component of the metric is estimated by calculating the strength of the gradient information transferred at edge positions. In order to identify pixels where edge component will be measured, a binary image, E_{valid} is constructed by combining edge images from visible, infrared and fused images, denoted, $E_{\text{L}_{\text{vis}}}$, E_{ir}, and $E_{\text{L}_{\text{fus}}}$, respectively: $E_{\text{valid}} = E_{\text{L}_{\text{vis}}} \vee E_{\text{ir}} \vee E_{\text{L}_{\text{fus}}}$. (The edge image is computed by applying Canny's edge detector to the lightness channel of each image.) We next compute the difference in gradients between the source and fused images at valid edge pixel positions: $d_{\text{V}_{\text{gradient}}} = \left|G_{\text{src}} - G_{\text{L}_{\text{fus}}}\right| \odot E_{\text{valid}}$, where $G_{\text{L}_{\text{fus}}}$ is the gradient image of the fused image. The edge component Rate_{GE} is given by the ratio:

$$\text{Rate}_{\text{GE}} = d_{\text{V}_{\text{gradient}}} \odot (G_{\text{src}})^{-1} \tag{3}$$

where, as before, the zero elements of G_{src} are replaced by unity to avoid division by zero.

The third component of the metric, a global contrast measure, is given by the ratio

$$\text{Rate}_{\text{GC}} = \frac{\left|GC_{\text{max}} - GC_{\text{L}_{\text{fus}}}\right|}{GC_{\text{max}}} \tag{4}$$

where $GC_{\text{max}} = \max\left\{GC_{\text{L}_{\text{vis}}}, GC_{\text{ir}}\right\}$ is the maximum between the global contrasts of the visual and infrared images, denoted $GC_{\text{L}_{\text{vis}}}$ and GC_{ir}, respectively. (The global contrast of each image is estimated as defined in [10]. For colour images the lightness channel is used for the estimation). Given any matrix A of dimensions $m \times n$, the average value of the entries of A is given by $\text{av}\{A\} = \frac{1}{mn}\sum_{i,j} a_{ij}$. Define the average error rate vector e_{rate} as follows:

$$e_{\text{rate}} = \begin{pmatrix} \text{av}\left\{\text{Rate}_{\text{hue}}\right\} \\ \text{av}\left\{\text{Rate}_{\text{GE}}\right\} \\ \text{Rate}_{\text{GC}} \end{pmatrix} \tag{5}$$

Finally, the proposed colour fusion metric (CFM) is computed as

$$\text{CFM} = \left|1 - \text{av}\left\{e_{\text{rate}}\right\}\right|. \tag{6}$$

It measures the transfer of information from the source images to the fused image. Observe that maximum information is transferred from the source to the fused image if the value of CFM is unity. It is possible for CFM to assume values greater than one, which usually implies artefacts and distortions have contributed to significant variations of the fused image compared to the source images.

III. Results and Discussions

A. Experimental Setup

For the comparative study, we have used five multi-scale decomposition-based fusion methods: the Laplacian pyramid (LP), non-subsampled contourlet transform (NSCT), NSCT with sparse representation (NSCT-SR), VIIF using *gradientlet* filter (GltFF), VIIF using a visual saliency map and weighted least square optimization (WLS-GF). For quantitative analysis of image fusion methods, the metrics used for comparison are the proposed performance colour-fusion metric (CFM) along with cross-entropy (CE), standard deviation (SD), spatial frequency (SF), and visual information fidelity (VIF) whose computational details are as in [9]. Observations from qualitative analysis are compared with the proposed fusion performance metric to test its efficiency. The experiments are conducted using MATLAB® 2023b on Intel® Core™ i9-14900K.

B. Datasets

Experimental Setup The following datasets are used in this paper because they have colour images with registered infrared images: (a) M^{3}FD Dataset: This is a Multi-scenario Multi-Modality benchmark dataset for the fusion of infrared and visible images for the purpose of object detection [11]. It contains registered visible and infrared images of resolution 1024×768 under various scenarios. We have used 290 images from this dataset. (b) LLVIP Dataset: This is a visible-infrared paired dataset for low-light vision for pedestrian detection [12]. It contains registered visible and infrared image pairs of resolution 1280×1024 for various low-light visual tasks. We have chosen 145 images from this dataset.

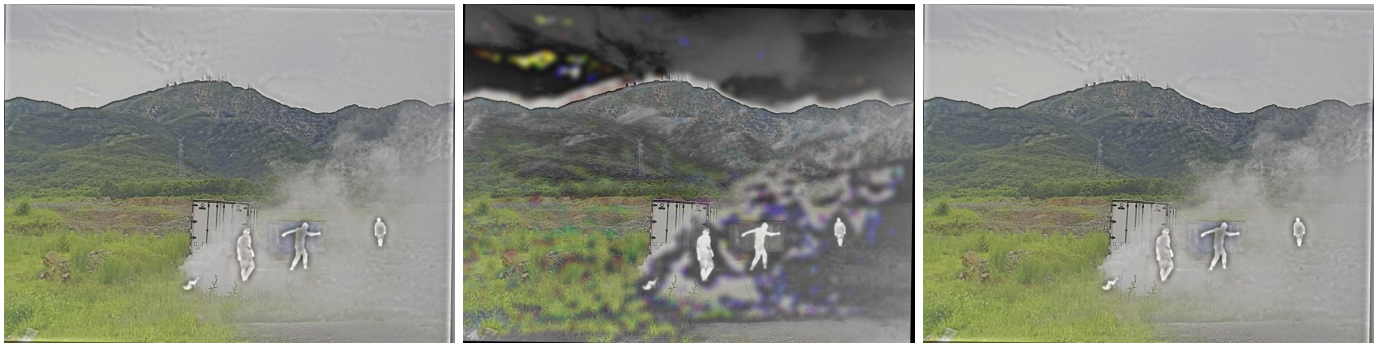
TABLE I: M^{3}FD Dataset. RGB: Values of Fusion Metrics.

Method	CE	SD	SF	VIF	CFM
GltFF	1.6025	32.8167	10.7045	0.4194	0.7957
LP	1.7506	32.1073	10.0829	0.3659	0.8010
NSCT	1.1481	36.0827	**14.5802**	0.4162	**0.8643**
NSCT-SR	**0.7150**	**46.1816**	14.5778	**0.6209**	0.5094
WLS-GF	1.8798	28.9401	13.643	0.4055	0.8159

TABLE II: M^{3}FD Dataset. LCH: Values of Fusion Metrics.

Method	CE	SD	SF	VIF	CFM
GltFF	1.6566	32.1187	10.0541	0.3988	0.8159
LP	1.3222	**37.5653**	**14.8248**	**0.4848**	0.8627
NSCT	**1.1469**	36.2141	14.5738	0.4137	**0.8720**
NSCT-SR	1.8275	34.1655	10.5287	0.4172	0.7511
WLS-GF	1.8303	29.0920	13.3569	0.403	0.8336

(a) Visible Image (b) Infrared Image



(c) RGB: NSCT (d) RGB: NSCT-SR (e) LCH: NSCT

Fig. 3: Example fusion of images from the M^3FD dataset.

C. Comparative Analysis and Discussions

As mentioned earlier, the performances of five fusion methods are analysed using the metrics CE, SD, SF and VIF. In addition, we study the efficacy of the proposed metric CFM. While a lower value of CE indicates better performance, higher values of other metrics are desirable. For the M^3FD dataset, the values of the five metrics are provided in Tables I and II, respectively, for the RBG colour space fusion framework (RGB fusion framework) and LCH colour space fusion framework (LCH fusion framework). From these values, it can be observed that the NSCT-SR method performed the best with respect to CE, SD and VIF metrics in the RGB fusion framework. In contrast, the CFM value indicates that this method results in the worst performance, which is visually indicated in Fig. 3. Here, the visible and infrared images are shown in Fig. 3(a) and Fig. 3(b), respectively. The fused results using the NSCT and NSCT-SR methods are depicted in Fig. 3(c) and Fig. 3(d), respectively. Evidently, from Fig. 3(d), the fused image using the NSCT-SR in the RGB colour space is quite unsatisfactory as it has considerable colour and edge artefacts. In contrast, the fused result with the NSCT method (see, Fig. 3(c)) indicates that pieces of information from both visible and infrared images are reasonably preserved. The SF and CFM metrics point to the NSCT method working best.

For the same dataset, the LP method provides the best performance according to SD, SF and VIF metrics in the LCH fusion framework (see Table II). Moreover, the CFM metric indicates acceptable performance. However, the CFM and CE metrics indicate that the NSCT method performs better in this framework. This is visually corroborated in Fig. 3(e). A comparison of Fig. 3(c) and Fig. 3(e) points to similar performances of the NSCT method in both frameworks. Nonetheless, the CFM values indicate that the latter framework works relatively better.

For the LLVIP dataset, the performances of the five methods are compared in Tables III and IV, respectively, in the RGB and LCH fusion frameworks. In the former framework, the NSCT-SR method yields the best performance. However, as seen in Fig. 4(d), this method introduces colour and edge artefacts, and the perceptual information obtained from the fused image can be quite different. In contrast to the NSCT-SR method, the LP method results in the best performance with respect to the CFM metric (Table III) which correlates better with qualitative performance as shown in Fig. 4(c). In the LCH fusion framework, all the metrics indicate that the LP method yields the best performance, as presented in Table IV. The fused image in Fig. 4(e) corroborates this observation. Moreover, the performances of the LP method in the RGB and LCH fusion frameworks are comparable, as indicated by Fig. 4(c) and Fig. 4(e). Further, the CFM values indicate that

(a) Visible Image (b) Infrared Image

(c) RGB: LP (d) RGB: NSCT-SR (e) LCH: LP

Fig. 4: Example fusion of images from the LLVIP dataset.

the LCH fusion framework works relatively better.

TABLE III: LLVIP Dataset. RGB: Values of Fusion Metrics.

Method	CE	SD	SF	VIF	CFM
GltFF	1.0409	39.8893	11.4403	0.5136	0.8340
LP	1.1048	35.5944	14.5201	0.6326	**0.8720**
NSCT	1.2408	34.0756	14.5393	0.5037	0.7324
NSCT-SR	**0.7882**	**45.7943**	**14.9712**	**0.7087**	0.5326
WLS-GF	1.1868	32.896	14.0497	0.4559	0.8029

TABLE IV: LLVIP Dataset. LCH: Values of Fusion Metrics.

Method	CE	SD	SF	VIF	CFM
GltFF	1.1236	37.1461	11.1693	0.4642	0.8487
LP	**0.9448**	**45.3922**	**14.8071**	**0.6412**	**0.8974**
NSCT	1.3093	32.8591	14.3244	0.4686	0.8884
NSCT-SR	3.7719	30.052	10.6831	0.3984	0.6926
WLS-GF	1.3219	30.7147	13.3345	0.4087	0.8174

For both the datasets, from Tables II and IV, the LP and NSCT methods using the LCH fusion framework, have resulted in fused images whose performance metrics are the best, as estimated by the metric CFM. From the results of experiments on the M^3FD dataset, it is observed that the performance of NSCT has not considerably improved when changing from RGB to LCH fusion framework. In contrast, in the LLVIP dataset, the fusion performance of the NSCT method has improved substantially. Although the NSCT-SR method improves its performance in the LCH fusion framework, it is still placed last based on the fusion performance evaluation using the CFM metric, which holds for tests on both datasets.

Evidently, the choice of the fusion method depends on the nature of the images in a dataset. For the M^3FD dataset, the NSCT method provides better visual perception relative to the other four methods in the context of this paper. This was contrary to the values of the existing metrics. In contrast, the proposed CFM metric appears to be more useful for human perception. Similarly, with the LLVIP dataset, the comparative analysis of the metrics results in the same observation that the CFM metric is a better choice. Moreover, the LP method performs best for low-light conditions amongst the five considered methods.

TABLE V: M^3FD Dataset. LCH: Component-wise CFM.

Method	Average	Hue	Edge	Contrast
GltFF	0.8159	0.9924	0.6819	0.7734
LP	0.8627	0.9957	0.7700	0.8225
NSCT	**0.8719**	**0.9972**	**0.7954**	**0.8231**
NSCT-SR	0.7510	0.9933	0.4642	0.7955
WLS-GF	0.8336	0.9908	0.7620	0.7478

(a) Visible Image (b) Infrared Image

(c) LCH: NSCT (d) LCH: WLS-GF

(e) RGB: NSCT-SR (f) LCH: NSCT-SR

Fig. 5: Fusion of Images from the M^3FD dataset.

In the previous analysis, the CFM metric is based on all the components of LCH. We emphasise that the CFM value of any component provides a similar conclusion. A component-wise study of the CFM metric is presented in Table V for a different example in the M^3FD dataset. Clearly, the NSCT method has performed the best for each component. The fused image from WLS-GF method has the lowest contrast as shown in Fig. 5(d) which is reflected well in the corresponding CFM value of 0.7478. Similarly, the NSCT-SR method results in edge artefacts as observed in Fig. 5(f), yielding the worst CFM value of 0.4642. Thus, the component-wise CFM is indicative of the resulting fused image.

Looking at fused images in Fig. 5(e) and Fig. 5(f), one can observe that only the colour artefacts have been removed. In contrast, edge artefacts have remained in the NSCT-SR fused image using the LCH fusion framework compared to the RGB fusion framework. Also, no such artefacts are present in Fig. 5(c), which is the fused result from the NSCT method using the LCH fusion framework. Adding a sparse representation-based method to construct the base layer of the fused image has introduced these artefacts.

IV. Conclusion

In this article, we have studied two frameworks for fusing visible images and infrared images. While one framework works directly in the RGB colour space, the colour image is first converted to LCH in the second framework. In the LCH fusion framework, the colour information is better transferred to the fused image when compared to the fused image resulting from the RGB fusion framework. Of the five methods compared here, the NSCT method works better for the M^3FD dataset. However, for low-light images, the LP method provides better-quality fused images. Finally, the proposed colour fusion metric is shown to better correlate with human perception when compared to other typically used metrics. The transference of colour, edge and contrast information can be evaluated with this metric.

Acknowledgment

This work was supported by the Aeronautics Research and Development Board (AR&DB), Defence Research and Development Organisation (DRDO), Ministry of Defence, Government of India, under Grant No. 2088.

References

[1] S. Singh, H. Singh, G. Bueno, O. Deniz, S. Singh, H. Monga, P. Hrisheekesha, and A. Pedraza, "A review of image fusion: Methods, applications and performance metrics," *Digital Signal Processing*, p. 104020, 2023.

[2] X. Zhang, P. Ye, and G. Xiao, "VIFB: A visible and infrared image fusion benchmark," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 104–105, 2020.

[3] Y. Shi, X. Jiang, and S. Li, "Fusion algorithm of UAV infrared image and visible image registration," *Soft Computing*, vol. 27, no. 2, pp. 1061–1073, 2023.

[4] Z. Shao, W. Wu, and S. Guo, "IHS-GTF: A fusion method for optical and synthetic aperture radar data," *Remote Sensing*, vol. 12, no. 17, p. 2796, 2020.

[5] S. Kalamkar *et al.*, "Multimodal image fusion: A systematic review," *Decision Analytics Journal*, p. 100327, 2023.

[6] X. Zhang, "Benchmarking and comparing multi-exposure image fusion algorithms," *Information Fusion*, vol. 74, pp. 111–131, 2021.

[7] 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, no. 1, pp. 94–109, 2011.

[8] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," *Aquatic Procedia*, vol. 4, pp. 133–142, 2015.

[9] X. Zhang, "Deep learning-based multi-focus image fusion: A survey and a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4819–4838, 2021.

[10] K. Matković, L. Neumann, A. Neumann, T. Psik, W. Purgathofer, *et al.*, "Global contrast factor: A new approach to image contrast," in *Proceedings of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, pp. 159–167, May 2005.

[11] J. Liu, X. Fan, Z. Huang, G. Wu, R. Liu, W. Zhong, and Z. Luo, "Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5802–5811, 2022.

[12] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou, "LLVIP: A visible-infrared paired dataset for low-light vision," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3496–3504, 2021.