Journal of Electrical Engineering, Vol. 76, No. 1, 2025, pp. 7-17

Bimodal and trimodal image fusion: A study of subjective scores and objective measures

Mohammed Zouaoui Laidouni [1], Boban P. Bondžulić *[1], Dimitrije M. Bujaković [1], Vladimir S. Petrović [2], Touati Adli [1], Milenko S. Andrić [1]

Thermal vision significantly enhances visibility under various environmental conditions. So, this paper presents a comprehensive study on the importance of thermal vision in improving image fusion human visual perception through subjective evaluation. The study focuses on the fusion of three imaging sensors commonly used in computer vision applications: long-wavelength infrared (LWIR), visible (VIS), and near-infrared (NIR). Four image fusion alternatives (LWIR+VIS, LWIR+NIR, NIR+VIS, and LWIR+NIR+VIS) are produced using a reliable deep learning approach and assessed using both subjective tests and objective metrics. The subjective evaluation is performed involving 15 military students and officers from the University of Defence in Belgrade, while objective assessment is elaborated using eight no-reference measures. Results indicate that fused images with thermal information show better visual performance than non-thermal based image fusion alternative (NIR+VIS). Moreover, LWIR+NIR+VIS and LWIR+NIR fused images provide similar visual appearance, demonstrating that the bimodal image fusion (LWIR+NIR) can be sufficient to produce a highly informative fused image. Additionally, the degree of agreement between subjective and objective scores is calculated. The simple edge intensity measure shows the highest degree of agreement, while the image entropy demonstrates the second-best score.

Keywords: thermal sensor, LWIR, NIR, VIS, image fusion, subjective quality assessment, objective quality assessment

1 Introduction

Thermal imaging sensors are sophisticated devices capable of detecting infrared radiation emitted by all objects with temperature above absolute zero [1]. They offer a unique opportunity that transcends the limitations of human vision and visible light imaging sensors. As a result, thermal sensors are increasingly employed across a diverse range of computer vision applications, such as object detection [2], tracking and surveillance [3, 4].

Considering the unique and valuable information provided by thermal sensors, their integration with other imaging sensors, such as visible (VIS) and near-infrared (NIR), becomes highly advantageous, in order to strengthen the overall visual capabilities. Through advanced image fusion techniques, the complementary information from these sensors can be combined to create composite images that significantly enhance scene understanding. This fusion is particularly beneficial in environments where VIS and NIR sensors alone cannot capture significant scene details, such as in low-light, foggy, or smoky conditions. By leveraging the distinct advantages of thermal imaging, fused images can improve the visibility and detection of objects, enhance contrast, providing a more robust and reliable representation for human vision [5, 6].

To generate as much as possible informative composite (fused) image, several notable fusion algorithms have been developed. They can be broadly categorized into two classes, task-specific image fusion [7-9] and general image fusion algorithms [10-12]. The first class focus is on combining two specific image modalities, typically LWIR+VIS or NIR+VIS. The second class consists of general image fusion algorithms that employ a unified model capable of fusing in a wider range of applications. These algorithms can handle various combinations such as LWIR+VIS, NIR+VIS, multi-exposure images, multi-focus images, and even medical image fusion, offering greater versatility across different applications. However, it is important to note that despite their ability to deal with various modalities, these algorithms typically fuse only two modalities at a time (bimodal fusion).

The current researches face two significant limitations. First, existing works primarily focus on fusing only two modalities for the same scene, (LWIR+VIS, or NIR+VIS), overlooking the potential to benefit of incorporating additional sensors that could provide complementary information. Second, there is a lack of subjective and objective assessments and comparison of fusion performance for different image combinations captured for the same scene (LWIR+VIS, LWIR+NIR, NIR+VIS, and LWIR+NIR+VIS). This gap in analysis

[1] Military Academy, University of Defence in Belgrade, Serbia
[2] Faculty of Technical Sciences, University of Novi Sad, Serbia
mohammedz.laidouni@gmail.com, * bondzulici@yahoo.com, adlitouati94@gmail.com, andricsmilenko@gmail.com, dimitrijebujakovic@gmail.com, vladimir.petrovic@uns.ac.rs

https://doi.org/10.2478/jee-2025-0002, Print (till 2015) ISSN 1335-3632, On-line ISSN 1339-309X

makes it challenging to determine the most effective sensor combination for image fusion task.

To address the limitations discussed above, this paper presents both, subjective and objective studies on the performance of fused images using various modality combinations. The subjective study highlights the importance of incorporating thermal channel in the fusion process to produce a more comprehensive and informative composite image.

This research makes the following key contributions:

- Multisensor image fusion analysis

  Our work expands the scope to three-modality (trimodal) fusion. The performance analysis and comparison of different sensor combinations (LWIR+VIS, LWIR+NIR, NIR+VIS, and the trimodal LWIR+NIR+VIS fusion) is performed.

- Comprehensive subjective study

  A subjective study is performed to evaluate the importance of different channels (sensors) in enhancing human visual perception.

- Subjective vs. objective assessment

  A comparison between subjective evaluations and objective assessment measures for image fusion is conducted to measure the degree of agreement between objective and subjective evaluations.

2 Database description

The database used in this research is created by collecting images from two different databases: TRICLOBS [13] and MOFA [14], which are obtained using multisensor surveillance cameras. Three sensors, including the thermal (LWIR), NIR, and VIS, are used in each database and the general characteristics of those sensors are presented in Tab. 1. The created dataset, along with VIS and NIR, contains LWIR images, which are assumed to provide enhanced information for improving human operator visual perception. In the aim of emphasizing different monitoring and surveillance scenarios under the outdoor conditions, the images have been chosen so that they include different objects. Therefore, the considered scenarios include: military and civilian individuals (stationary, walking, or running, carrying various objects, and groups of soldiers), as well as civilian and military vehicles.

The database contains 192 image triplets, each of them describes same scene. To demonstrate the effectiveness of thermal information in image fusion, both trimodal and bimodal fused images are produced (LWIR+VIS, LWIR+NIR, NIR+VIS, and LWIR+NIR+VIS) for comparative analysis. The database is split, using 162 triplets for training while the remaining 30 are used for testing, employing both subjective and objective assessment methods.

Table 1. General characteristics of the TRICLOBS and MOFA sensors

Sensor	Characteristics	TRICLOBS	MOFA
LWIR	Detector	microbolometer	Huajingkang Infrared K26E19
	Spectral range	8-14 μm	8-14 μm
	Resolution	480×640	440×600
NIR	Detector	enhanced CMOS	Daheng Imaging MER-502-79UCM
	Spectral range	0.7-1 μm	0.78-1 μm
	Resolution	480×640	440×600
VIS	Detector	enhanced CMOS	Daheng Imaging MER-502-79U3C
	Spectral range	400-700 nm	400-700 nm
	Resolution	480×640	440×600

3 Multisensor image fusion architecture

In recent years, a growing number of image fusion methods have been proposed, including: 1) traditional approaches, such as transform domain-based methods, saliency-based methods, sparse representation-based methods and hybrid methods [6], and 2) methods based on deep learning (DL), such as CNN-based methods, GAN-based methods, transformer-based methods and ConvNeXt-based methods [7, 9]. This article focuses on DL methods due to their superior performance and ability to automatically learn complex features from data, which makes them particularly well-suited for enhancing image fusion tasks [9].

While CNNs offer simplicity but lower accuracy, and transformers provide high accuracy at the cost of increased complexity, this study employs ConvNeXt, a novel architecture that combines the benefits of both approaches. ConvNeXt effectively combines the simplicity and efficiency of CNNs with the powerful feature learning capabilities of transformers, making it a promising approach for achieving enhanced image fusion results. It is a neural network architecture with a large-kernel size convolution, constructed entirely from the standard ResNet, which adopts several modern techniques inspired by transformer architecture, such as layer normalization, Gaussian error linear unit activetion, and depth-wise convolutions [15].

Figure 1 presents the overall diagram of the used multisensor image fusion architecture based on ConvNeXt layers.

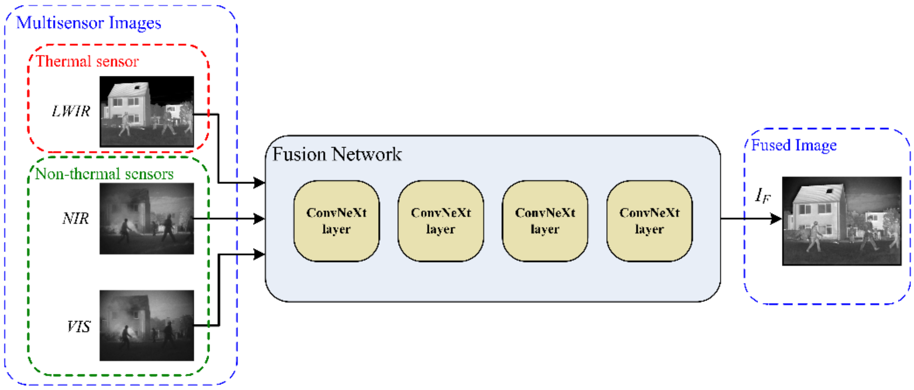

Fig. 1. The overall diagram of multisensor image fusion architecture

The proposed architecture aims to generate the fused image I_F via merging thermal (LWIR) and non-thermal images (NIR and VIS). As illustrated in Fig. 1 the architecture is structured into three main parts: multisensor image inputs, the fusion network, and the fused image output.

The input images are fed into the fusion network. The fusion network uses several interconnected ConvNeXt layers to integrate complementary information from the input images. This architecture is designed to preserve both the thermal information captured by the thermal sensor and the detailed texture present in the non-thermal images, ensuring a comprehensive fusion of the multisensor inputs. Finally, the fused image is produced as the output.

4 Experiments and analyses

In this section, an experimental validation is conducted to evaluate the effectiveness of the thermal radiation information in enhancing the human visual perception. At the beginning, the fusion of thermal image with NIR and VIS images is conducted to combine the thermal radiation information with rich scene details. Next, the thermal information effect on image fusion is studied via subjective assessment. Finally, the evaluation of the degree of agreement between the subjective and objective image fusion scores is performed.

4.1 Trimodal image fusion

In this section, the fusion of the three sensor images is conducted. Due to the lack of DL methods designed for handling all three channels simultaneously, our ConvNeXt-based approach will be compared against three traditional methods that do not rely on DL. These methods include Laplacian Pyramid (LP) [16], Latent Low-Rank Representation (LatLRR), and Guided Filtering (GFF) [17].

To objectively assess the performance of the three-sensors image fusion, three metrics have been used, including average gradient (AG), entropy (EN) [18], and Naturalness Image Quality Evaluator (NIQE) [19]. Tab. 2. presents the objective results, where the best results are highlighted.

Table 2. Objective comparison of trimodal image fusion

Method	LP	GFF	LatLRR	Our
AG	3.0559	3.9580	3.1904	**4.5736**
EN	6.7962	7.0918	7.0002	**7.4955**
NIQE	4.3335	4.9291	4.1738	**3.3600**

From Tab. 2, it is observed that our ConvNeXt-based method outperforms the other three methods (LP, GFF, and LatLRR) across all three quantitative metrics. It shows significant improvements in gradient and entropy (higher is better), indicating better preservation of structural details and information content. The lower NIQE score also suggests improved perceptual quality.

Fig. 2. Visual comparison of trimodal image fusion

Figure 2 illustrates two challenging in image fusion scenarios: a scene under low light conditions and another obscured by smoke. As one can see, LP, LatLRR, and GFF methods struggle to adequately extract thermal radiation information from the thermal sensor in both scenarios, resulting in fused images with low contrast. Otherwise, our approach exhibits superior performance in extracting thermal radiation data, effectively integrating visual information from other sensors, and preserving high contrast in the fused output. This enhanced capability allows for more effective visualization and interpretation of complex scenes under adverse conditions. Therefore, in the further part of the paper, ConvNeXt-based trained models are used for both trimodal and bimodal image fusion.

4.2 Thermal information effect in image fusion

In order to evaluate the visual effect of thermal information on the fusion performance, four fused image alternatives (LWIR+VIS, LWIR+NIR, NIR+VIS, and LWIR+NIR+VIS) are produced for testing. Both subjective and objective assessments are conducted to evaluate the fusion results. The subjective evaluation is performed by observers at the University of Defence in Belgrade to compare the quality of fused images. As observers, 15 military students and officers are randomly selected to ensure a different level of familiarization with image fusion. They were asked to compare the four fused images and rank them with a ranking score (RS) from "1" to "4", where "1" corresponds to the image with the best fused results, and "4" to the image with the worst fused results.

The objective evaluation is conducted using eight no-reference metrics, including average gradient (AG), spatial frequency (SF), entropy (EN), variance (VAR), edge intensity (EI) [18], NIQE [19], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [20], and Perception based Image Quality Evaluator (PIQE) [21]. The results of the subjective and objective studies are discussed in the next subsections.

4.2.1 Subjective image fusion quality assessment

Figure 3 presents the gathered subjective results of the four image fusion alternatives LWIR+VIS, LWIR+NIR, NIR+VIS, and LWIR+NIR+VIS (ALL), utilizing mean ranking score (MRS), and standard deviation (MRS±σ).

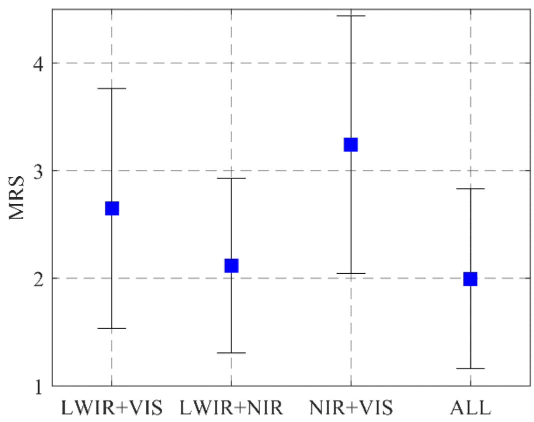

Fig. 3. Image fusion subjective results through mean ranking scores (MRS) and their standard deviations (MRS±σ)

The mean ranking scores in Fig. 3 show the overall performance of the four image fusion alternatives. It is observed that NIR+VIS fusion that excludes thermal information, exhibits the highest MRS (the worst fusion results), as well as the highest standard deviation. In contrast, the alternatives incorporating thermal information demonstrate superior performance, as indicated by lower MRS values, showing the best results for the combination of all sensors. This analysis supports the importance of including thermal (LWIR) information in image fusion for enhanced visual performance, with the best results achieved when combining all available sensors. Also, an excellent alternative to trimodal is bimodal LWIR+NIR image fusion.

Figure 4 presents the MRS evaluation for each observer. It is noticed that the individual preferences of each observer tend to choose the LWIR+NIR+VIS, and LWIR+NIR with less degree, as the best image fusion results. Furthermore, the NIR+VIS fusion is chosen as the worst alternative by almost all subjects except for, observers 4, 8, 9, 12 and 15. That emphasize the importance of thermal information captured by LWIR sensors in individual human visibility perception.

Figure 5 shows MRS results by scene for all observers, and two observations can be noted. Firstly, NIR+VIS image fusion shows the worst subjective results across almost all scenes. Moreover, it is selected as the worst method for scenes 9, 14, 17, 22, 24, 29 and 30 by all candidates. Secondly, LWIR+NIR fusion shows a very competitive MRS performance with LWIR+NIR+VIS fusion. Furthermore, it is observed that LWIR+VIS fusion is chosen by all subjects as the best results for scene 10. Again, this highlights the importance of incorporating thermal (LWIR) sensors in the fusion process to enhance the visual quality across different scenes.

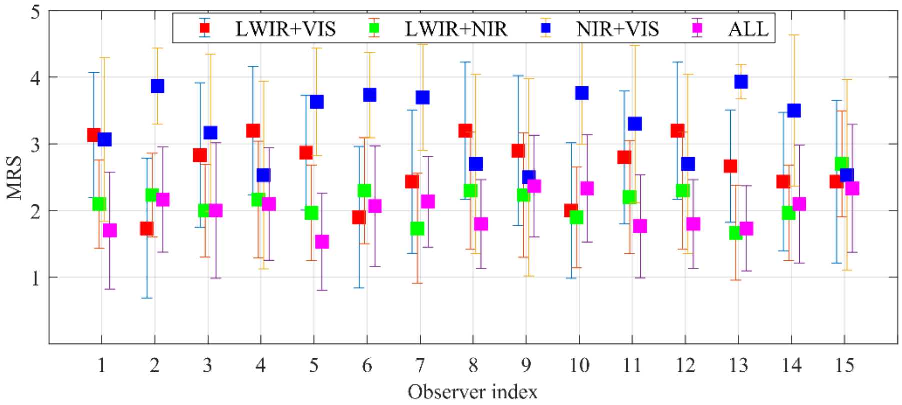

Fig. 4. MRS with standard deviations for each observer

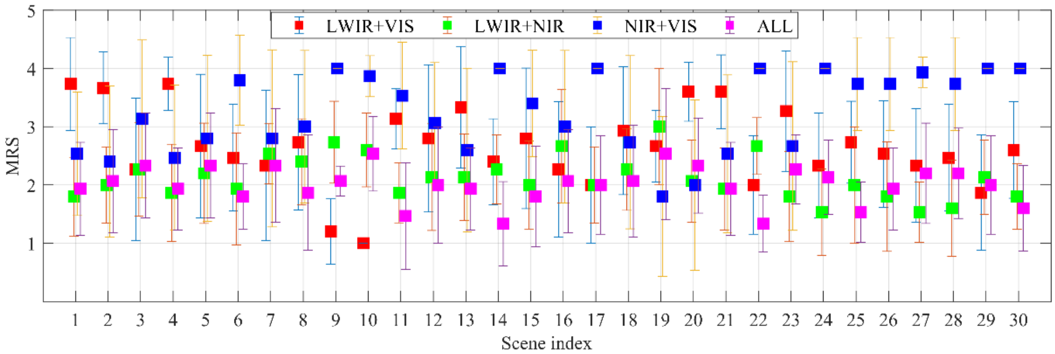

Fig. 5. MRS with standard deviations for each scene

4.2.2 Examples of bimodal and trimodal image fusion

Figure 6 presents two LWIR, NIR and VIS test image triplets (scenes 5 and 27) and the corresponding bimodal and trimodal fused images. All image fusion alternatives can relatively exhibit a good visual performance. However, there is some drawbacks noted by a visual comparison. NIR+VIS fused images cannot capture the scene details in case of dim and shows poor contrast, resulting in obscured objects. The incorporation of thermal sensor in image fusion process significantly improves the quality of fused images by enhancing the contrast, resulting in emphasizing objects (see soldiers, people and cars). LWIR+VIS fused images show a good contrast but loses some details of the scene (see the texture in the building, street, and clouds in the sky). LWIR+NIR fused images show better contrast and capability in describing the scene details, but loses some texture (see the walking street). Finally, LWIR+NIR+VIS fused images maintain the thermal radiation resulting in good contrast, and also captures the scene details from NIR and VIS images.

Created dataset, the results of subjective tests, ConvNeXt-based implementation code and pre-trained weights are available on Mendeley Data link [22].

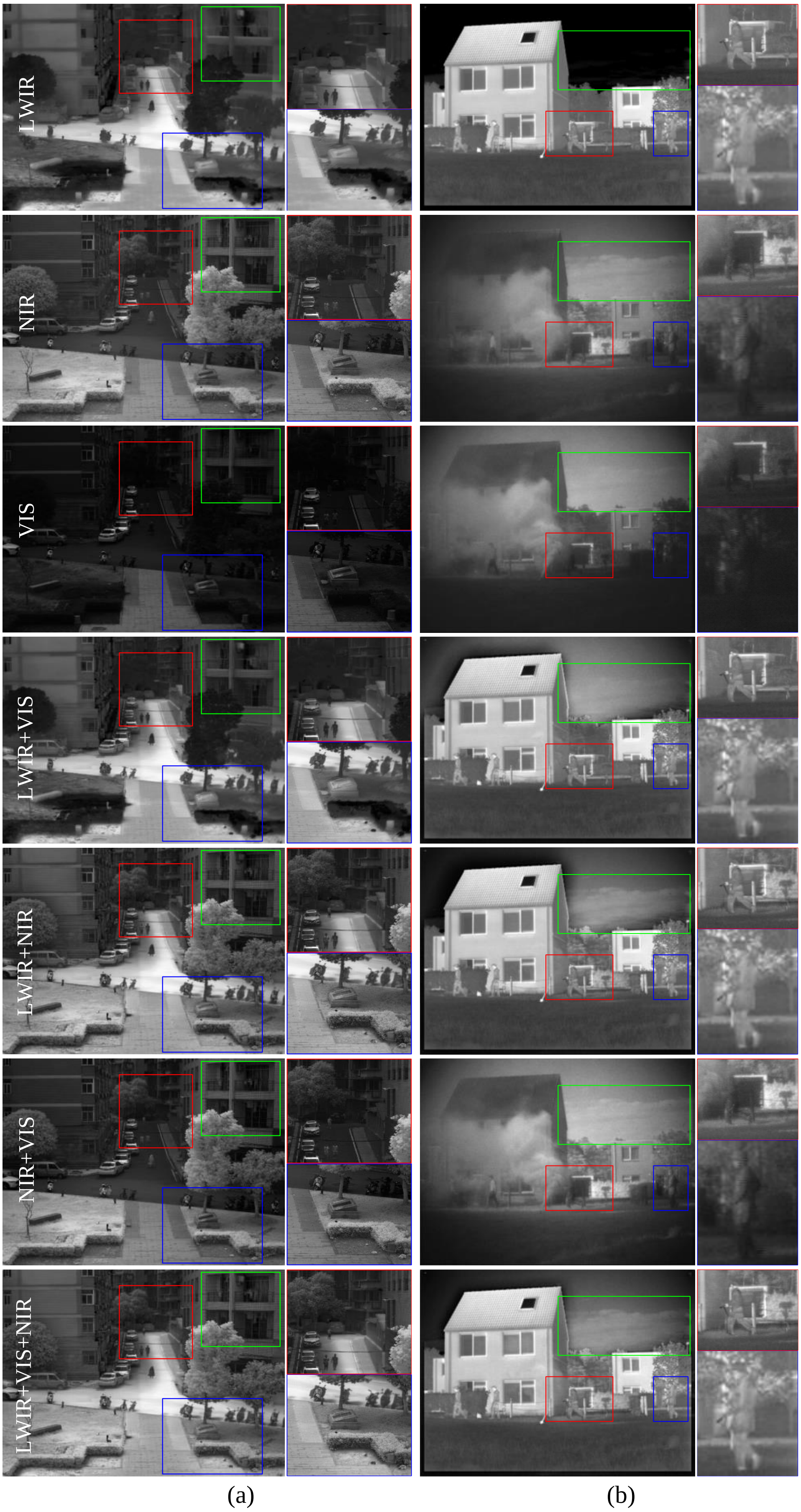

Fig. 6. Bimodal and trimodal image fusion examples: (a) scene 5, and (b) scene 27

5 Objective vs subjective quality assessment

In this section, the obtained results between the objective and subjective quality assessment for the four image fusion alternatives are discussed. Therefore, the objective assessment between bimodal and trimodal image fusion alternatives is presented in Tab. 3, while Fig. 7 depicts the scatter plots of subjective MRS against the eight no-reference objective metrics mean scores.

Table 3. Bimodal and trimodal image fusion performance comparison

Fused sensors	AG	SF	EN	VAR	EI	NIQE	BRISQUE	PIQE
LWIR+VIS	4.0043	12.3713	7.4085	54.9334	41.8896	3.1337	25.7947	32.8673
LWIR+NIR	4.5307	13.6844	7.4708	55.8054	46.4951	3.0935	24.9049	31.3570
NIR+VIS	3.8143	11.0811	7.1197	43.6548	36.7359	5.0090	23.1004	22.0794
ALL	4.5736	13.4720	7.4955	55.1148	46.9963	3.3600	23.2153	29.3230

From Tab. 3, it can be concluded that the three simple no-reference measures AG, EN, and EI selected trimodal image fusion as the best solution, while according to them a good alternative is bimodal LWIR+NIR image fusion, and this fully agrees with the subjective test results (see Fig. 3). According to SF and VAR objective measures, these two image fusion alternatives are also the best, but in reverse order. The reliability of AG, EN, and EI objective measures is further confirmed in Fig. 7, where, in addition to the ranking that is fully consistent with the results of the subjective tests, it can be concluded that the mutual distances of the mean objective scores also agree with the subjective MRS distances (the mean LWIR+VIS+NIR and NIR+VIS scores are closer than the other scores). Reliable perceptual objective image quality assessment measures NIQE, BRISQUE, and PIQE do not match well with MRS. This misalignment can be explained by the fact that these metrics are primarily designed for visible light (VIS) images, thus lacking effectiveness when applied to multimodal image fusion quality assessment. Additionally, predefined models used in NIQE and BRISQUE measures are obtained using images of natural (non-thermal) scenes. Also, the BRISQUE and PIQE measures select bimodal NIR+VIS as the best image fusion approach, i.e. a method that does not use fusion of thermal information.

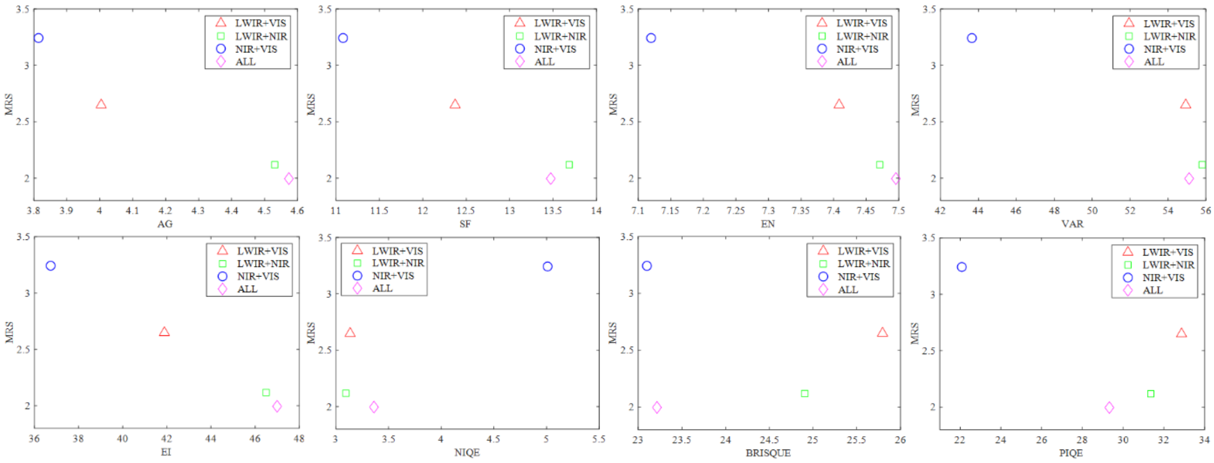

Fig. 7. Scatter plots of the MRS against the objective image fusion metrics

Furthermore, to measure the degree of agreement between the subjective and objective evaluations, the results of the eight no-reference metrics are used to rank the four fused image alternatives for each scene, from “1” to “4”, where “1” corresponds to the image with the best fusion results, and “4” to the image with the worst fusion results. After those two standard measures are used to calculate the error (distance) between the objective and the subjective ranking scores, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The obtained results are presented in Tab. 4.

Table 4. Comparison between the subjective and objective ranking scores

Metric	AG	SF	EN	VAR	EI	NIQE	BRISQUE	PIQE
MAE	0.9000	1.0667	0.7833	0.9333	0.7500	0.9000	1.1500	1.2667
RMSE	1.1123	1.3151	0.9922	1.1516	0.9258	1.1147	1.3964	1.5623

Based on the obtained results, EI exhibits the best agreement with the subjective results in term of both MAE and RMSE. Moreover, it is noticed that EN demonstrate the second-best agreement with subjective results, while AG and NIQE show the third best agreement. Given that EI performs best, it might be worth considering edge intensity as a primary no-reference metric for assessing the performance of multimodal image fusion (LWIR, NIR and VIS).

However, it is important to emphasize that although EI provides the best results among the tested metrics, both at the level of the complete database (Fig. 7), and through the comparison of ranks at the level of individual scenes (Tab. 4), the relatively high value of the rank differences, MAE, indicates that there is a need for further improvements in the area of multimodal image fusion quality assessment. A special challenge here is the quality assessment when the fused image is obtained with a different number of source images or based on images of different modalities.

6 Conclusion

This paper presents a subjective and objective studies on the performance of four image fusion alternatives, bimodal (LWIR+VIS, LWIR+NIR, NIR+VIS) and trimodal (LWIR+NIR+VIS). A reliable deep learning ConvNeXt-based trained network is used for image fusion alternatives. The study emphasizes the significance of the thermal information in enhancing human visibility through fusion process.

The subjective assessment was conducted by 15 military students and officers at the University of Defence in Belgrade. Three analyses were performed to discuss and evaluate the obtained ranking scores: mean ranking scores by image fusion alternatives, mean ranking scores by observers, and mean ranking scores by scenes. Based on subjective quality assessments, it was concluded that fused images with thermal information show better performance than the bimodal approach that does not consider the thermal channel. Moreover, on the created dataset, LWIR+NIR and LWIR+NIR+VIS image fusion alternatives show a close performance, indicating that the bimodal image fusion (LWIR+NIR) can be sufficient to produce a highly visually informative fused image.

The objective assessment was elaborated using eight no-reference metrics. In addition, the error between the objective and the subjective rankings was calculated, to measure the degree of their agreement. Results indicate that edge intensity shows the best agreement, while entropy provides the second-best score. This dual approach provides a more understanding of image fusion performance assessment, bridging the gap between computational measures and human perception by analyzing the agreement between subjective and objective assessments. Thus, we decided to make the created image database, along with the results of subjective tests, publicly available to the research community.

Considering the agreement between subjective and objective results, it is necessary to develop new or improved well-performing objective metrics that might provide a more robust and comprehensive image fusion quality assessment across various number of sensors, various modalities and fusion techniques, which will be the subject of our future work.

Acknowledgment

This research has been a part of Project No. VA/TT/1/25-27 supported by the Ministry of Defence, Republic of Serbia.

References

[1] O. Zelmati, B. Bondzulic, B. Pavlovic, I. Tot, and S. Merrouche, "Study of subjective and objective quality assessment of infrared compressed images", *Journal of Electrical Engineering*, vol. 73, no. 2, pp. 73–87, doi: 10.2478/jee-2022-0011, 2022.

[2] D. Peric, B. Livada, M. Peric, and S. Vujic, "Thermal imager range: Predictions, expectations, and reality", *Sensors*, vol. 19, no. 15, pp. 3313, doi: 10.3390/s19153313, 2019.

[3] R. Gade, and T. B. Moeslund, "Thermal cameras and applications: A survey", *Machine Vision and Applications*, vol. 25, no. 1, pp. 245-262, doi: 10.1007/s00138-013-0570-5, 2014.

[4] D. L. Hickman, and S. J. Shepperd, "Image fusion systems for surveillance applications: Design options and constraints for a tri-band camera", *Proceedings SPIE Infrared Technology and Applications XLVII*, pp. 310-328, doi: 10.1117/12.2584984, 2021.

[5] C. Jiang, H. Ren, H. Yang, H. Huo, P. Zhu, Z. Yao, J. Li, M. Sun, and S. Yang, "M2FNet: Multi-modal fusion network for object detection from visible and thermal infrared images", *International Journal of Applied Earth Observation and Geoinformation*, vol. 130, pp. 103918, doi: 10.1016/j.jag.2024.103918, 2024.

[6] R. Li, M. Zhou, D. Zhang, Y. Yan, and Q. Huo, "A survey of multi-source image fusion", *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 18573-18605, doi: 10.1007/s11042-023-16071-9, 2024.

[7] Y. Hua, W. Xu, and Y. Ai, "A residual ConvNeXt-based network for visible and infrared image fusion", *4th International Conference on Electronic Communication and Artificial Intelligence (ICECAI)*, Guangzhou, China, May 12-14, pp. 370-376, doi: 10.1109/ICECAI58670.2023.10176540, 2023.

[8] R. Soroush, Y. Baleghi, "NIR/RGB image fusion for scene classification using deep neural networks", *The Visual Computer*, vol. 39, no. 7, pp. 2725-2739, doi: 10.1007/s00371-022-02488-0, 2023.

[9] X. Zhang, and Y. Demiris, "Visible and infrared image fusion using deep learning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no 8, pp. 10535-10554, doi: 10.1109/TPAMI.2023.3261282, 2023.

[10] 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, doi: 10.1016/j.inffus.2014.09.004, 2015.

[11] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: A unified unsupervised image fusion network", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 502-518, doi: 10.1109/TPAMI.2020.3012548, 2022.

[12] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma, "SwinFusion: Cross-domain long-range learning for general image fusion via Swin Transformer", *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1200-1217, doi 10.1109/JAS.2022.105686, 2022.

[13] A. Toet, M. A. Hogervorst, and A. R. Pinkus, "The TRICLOBS dynamic multi-band image data set for the development and evaluation of image fusion methods", *PLoS One*, vol. 11, no. 12, pp. 0165016, doi: 10.1371/journal.pone.0165016, 2016.

[14] K. Xiao, X. Kang, H. Liu, and P. Duan, "MOFA: A novel dataset for multi-modal image fusion applications", *Information Fusion*, vol. 96, pp. 144-155, doi: 10.1016/j.inffus.2023.03.012, 2023.

[15] Z. Liu, H. Mao, C.Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s, *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, USA, June 18-24, pp. 11966-11976, doi: 10.1109/CVPR52688.2022.01167, 2022.

[16] P. J. Burt, and E. H. Adelson, "The Laplacian pyramid as a compact image code", *IEEE Transactions on Communication*, vol. 31, no. 4, pp. 532-540, doi: 10.1016/B978-0-08-051581-6.50065-9, 1983.

[17] X. Zhang, P. Ye, and G. Xiao, "VIFB: A visible and infrared image fusion benchmark", *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, June 14-19, pp. 468-478, doi: 10.1109/CVPRW50498.2020.00060, 2020.

[18] S. Singh, H. Singh, G. Bueno, O. Deniz, S Singh, H. Monga, P. N. Hrisheekesha, and A. Pedraza, "A review of image fusion: Methods, applications and performance metrics", *Digital Signal Processing*, vol. 137, pp. 104020, doi: 10.1016/j.dsp.2023.104020, 2023.

[19] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer", *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209-212, doi: 10.1109/LSP.2012.2227726, 2013.

[20] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain", *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695-4708, doi: 10.1109/TIP.2012.2214050, 2012.

[21] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features", *Twenty First National Conference on Communications (NCC)*, Mumbai, India, 27 February - 01 March, pp. 1-6, doi: 10.1109/NCC.2015.7084843, 2015.

[22] Z. M. Laidouni, B. Bondžulić, D. Bujaković, V. Petrović, T. Adli, and M. Andrić, "BTIFD: Bimodal and trimodal image fusion database", *Mendeley Data*, V1, available at: https://data.mendeley.com/datasets/btnws5tbcm/1, 2024.

Mohammed Zouaoui Laidouni received the Undergraduate Scientific degree at National Preparatory School for Engineering Studies, Algiers, Algeria in 2014 and M.Sc. degree in telecommunication engineering from Ecole Military Polytechnic, Algiers, Algeria in 2017. He is a doctoral student at Military Academy, University of Defence in Belgrade. His research interests are image fusion and image quality assessment. He has published 4 papers in national and international conferences and journals.

Boban Bondžulić received the B.Sc. degree in electrical engineering from Military Technical Academy, Serbia in 2000, M.Sc. degree at the School of Electrical Engineering, University of Belgrade, Serbia in 2005 and Ph.D. degree from Faculty of Technical Sciences, University of Novi Sad, Serbia in 2016. He is an Associate Professor at the Department of Telecommunications and Informatics, Military Academy, University of Defence in Belgrade. His research interests include information fusion, image and video quality evaluation, detection and tracking of moving objects, pattern recognition. He has published more than 130 papers in national and international conferences and journals.

Dimitrije Bujaković received the B.Sc. degree in electrical engineering from Military Technical Academy, Serbia in 2004, M.Sc. degree at the School of Electrical Engineering, University of Belgrade, Serbia in 2008 and Ph.D. degree from the School of Electrical Engineering, University of Belgrade, Serbia in 2016. He is an Associate Professor at the Department of Military Electrical Engineering, Military Academy, University of

Defence in Belgrade. His research interests include pattern recognition and methods for signals analysis and digital signal processing. He has published more than 50 papers in national and international conferences and journals.

Vladimir Petrović, graduated with MEng and PhD in electronic and electrical engineering and digital image processing at the University of Manchester UK in 1997 and 2001. Until 2016 he was with the Imaging Science Biomedical Engineering Group, University of Manchester UK, and since 2009 he has been a Professor of Signal Processing at the University of Novi Sad, Serbia. His research interests include medical image analysis and processing, image and video quality metrics, machine vision, information and image fusion and have resulted in over 60 publications in these areas.

Touati Adli received the Undergraduate Scientific degree at National Preparatory School for Engineering Studies, Algiers, Algeria in 2016 and M.Sc. degree in computer science engineering from Ecole Military Polytechnic, Algiers, Algeria in 2019. He is a doctoral student at Military Academy, University of Defence in Belgrade. His research interests include machine/deep learning and computer vision. He has published 4 papers in national and international conferences and journals.

Milenko Andrić received the B.Sc. degree in electrical engineering from Military Technical Academy, Serbia in 1995, M.Sc. degree at the School of Electrical Engineering, University of Belgrade, Serbia in 2001 and Ph.D. degree from the Military Academy, Serbia in 2006. He is a full-time professor at the Department of Military Electrical Engineering, Military Academy, University of Defence in Belgrade. His main research interests are in the fields of stochastically process in telecommunication and radar engineering, pattern recognition, methods for signals analysis and digital signal processing. He has published more than 110 papers in national and international conferences and journals.

Received 15 September 2024