

ENHANCING IMAGES IN POOR LIGHTING CONDITIONS THROUGH FUSION OF OPTICAL AND THERMAL CAMERA DATA

*Kostiantyn Abramenzkov, Dmytro Hryshchak**Dnipro University of Technology, 19, Dmytra Yavornytskoho Ave., Dnipro, 49005, Ukraine.**Authors' e-mails: Abramenzkov.K.M@nmu.one, Hryshchak.D.D@nmu.one*<https://doi.org/10.23939/acps2025.02.229>*Submitted on 30.09.2025*

© Abramenzkov K., Hryshchak D., 2025

Abstract: The goal of the article is to provide a methodology of improving images quality in low-light conditions through fusion of data received from telecamera and thermal camera. Data from thermal camera uses for compensation of significant illumination reduction in poor lighting conditions and allow keep required level of information. Proposed method establishes dynamic regulation of fusion coefficients depending on brightness level to minimize artifacts, increase edge sharpness, and improve object detectability. Developed model enables investigation of the influence of algorithmic parameters on key quality indicators, particularly PSNR, SSIM and target detection metrics. It has been shown that implementation of adaptive multimodal fusion principles ensures an increase in structural similarity by 15-20% and improvement in target detection accuracy in dark scenes by over 25% compared to using individual channels.

Index terms: multimodal fusion, thermal camera, telecamera, low lightning, computer vision, deep learning.

I. INTRODUCTION

Computer vision systems in low-light conditions have long been applied in security, monitoring, and navigation tasks, but their effectiveness is significantly limited by the physical characteristics of sensors. Even modern television cameras capable of operating at minimal illumination remain sensitive to increased noise levels, loss of contrast, and reduced detail. This is particularly critical in night or foggy scenes, when visual information loses a significant portion of its informativeness. Under such conditions, any instability in lighting, defocus, or glare quickly transforms into detection or object identification errors. In the practice of optical system applications, the problem of target detection in darkness is often solved either by increasing sensor sensitivity or by additional scene illumination. However, both approaches have limitations: the first increases noise levels, the second unmasks the system or alters the natural appearance of the scene. Thermal cameras, in turn, provide stable operation in darkness but produce images with limited detail and sometimes excessive thermal intensity, which complicates interpretation. Recent years have demonstrated a gradual shift in approach: instead of using a single image channel, solutions for combining data from different sensors are emerging – primarily television and thermal cameras.

However, existing fusion algorithms often work only under idealized conditions, do not account for variable scene factors, and do not adapt to specific tasks. The problem is not a lack of technologies – the problem is that they do not dynamically adjust to changes in lighting, spectral characteristics, and shooting conditions.

Therefore, the task of creating multimodal fusion approaches that would account not only for differences in spectral ranges and image geometry, but also for the dynamics of changing observation conditions is currently relevant: how scene intensity and contrast change, how atmospheric factors influence, how the weight of each channel is adjusted depending on local frame characteristics. This task lies at the intersection of optical engineering, signal processing, and deep learning methods. Without solving it, even the most modern systems will remain partially empirical and insufficiently reliable under complex operating conditions.

II. LITERATURE REVIEW AND PROBLEM STATEMENT

Multimodal fusion of images from television and thermal cameras is one of the key directions in the development of modern computer vision systems for operation in low-light conditions. Within this approach, various algorithmic and architectural solutions are applied from classical spatial-frequency processing methods to deep neural networks capable of adaptively adjusting channel weighting coefficients depending on scene characteristics [1]. In work [2], an enhancement method using depth information is proposed, which allows improving contrast and object detail in complex scenes. Special attention is paid to algorithms focused on improving detection efficiency. Study [3] presents the real-time RMF-ED algorithm, which combines data from two channels, reducing artifacts and increasing edge sharpness in darkness. A similar effect is achieved in work [4], where a transformer architecture is used for multispectral pedestrian detection, enabling increased recognition accuracy under limited lighting conditions. Meanwhile, research [5] examines the influence of non-uniform thermal distribution on fusion results and proposes an algorithm to compensate for this effect.

Deep learning methods are actively applied for integrating data from different sensors. In [6], a mul-

tispectral DNN model for object detection in darkness is created, and work [7] demonstrates the effectiveness of thermal modality in recovering details under low lighting. Additionally, [8] presents the NOT-156 dataset, which combines low-light images and thermal frames for developing nighttime tracking algorithms.

The creation of specialized benchmarks and lightweight models is a separate promising direction. Specifically, work [9] presents the Thermal-Aware LIE model, optimized for real-world conditions, and [10] proposes a preprocessing method that integrates infrared and visual channels to improve night vision quality.

A separate group consists of studies dedicated to geometric and spectral correction of multi-channel images. In [11], a technology for geometric and spectral correction of optical-electronic space images is described, ensuring channel consistency before further analysis. Work [12] examines pansharpening technology, which allows combining multispectral and panchromatic images to increase detail; this approach is logically similar to the task of multimodal fusion of visible and thermal channels. Similar multispectral data processing techniques are also investigated in study [13], which proposes an algorithm for preprocessing space images based on integration of multiple spectral ranges. In this work, an information technology for increasing the informativeness of multi-channel data using packet wavelet transforms is developed, demonstrating the potential of such approaches for adaptive quality improvement. Study [14] substantiates a hyperspherical transformation method that enables increasing spatial resolution of multi-channel aerospace images.

Thus, a review of recent works shows a wide spectrum of approaches – from classical spatial correction methods to deep models and transformer architectures. However, the common challenge remains the problem of algorithm adaptability: most of them work well only under idealized conditions, while in real dynamic scenes (fog, nighttime conditions, non-uniform lighting) their effectiveness decreases. This emphasizes the need for developing hybrid solutions that would combine the accuracy of classical methods with the flexibility of deep learning and could operate in real time.

III. SCOPE OF WORK AND OBJECTIVES

The research objective is to develop and substantiate a comprehensive approach to improving image quality in low-light conditions through multimodal fusion of data from television and thermal cameras, which involves the integration of deep learning methods, spatial-frequency analysis, and adaptive regulation of channel weighting coefficients to ensure stable detail preservation, artifact reduction, and improved object detection accuracy in real time.

During modeling and experimental verification, it was established that the application of multimodal fusion of television and thermal images enables a significant increase in scene informativeness under low-light conditions. This is explained by the fact that each channel

carries its own type of information: the television channel with high detail reproduces textural features but loses quality in the presence of noise and reduced contrast, while the thermal channel provides stable silhouette reproduction regardless of lighting but lacks sufficient spatial detail. Their combination in a single information field allows compensating for the weaknesses of each sensor and achieving a synergistic effect.

Preliminary sensor calibration proved to be a critically important stage, as it enabled reduction of geometric errors and increased accuracy of spatial correspondence between channels. The absence of such calibration would lead to artifacts during frame registration and degradation of subsequent fusion quality. The next step was bicubic interpolation of the lower-resolution image, which made it possible to bring both streams to a unified scale without significant loss of contrast and sharpness. This created a foundation for consistent analysis of pixel relationships.

To prevent color distortions, a transition to HSV space was applied, where regulation of channel weighting coefficients became simpler and more controllable. In the practical implementation of the pipeline, several key procedures were provided: gentle detail enhancement in the HD frame using unsharp mask (kernel 3×3 , $\sigma=1.0$, amount=0.8), ISO image denoising through bilateral filter ($d=9$; $\sigma_{\text{color}}=75$; $\sigma_{\text{space}}=75$), subsequent application of exposure-fusion with weights close to 0.8/0.2 in favor of the visible channel, and final local contrast enhancement through CLAHE on the L-channel (clip limit 1.5). Under conditions of pronounced lighting non-uniformity, an adaptive branch was introduced: the noise suppression strength for the ISO frame and the sharpening degree for HD were automatically adjusted based on local noise estimates and gradient response. Additionally, in dark areas of the scene, local brightness correction with contrast limiting was applied, after which the results were coordinated using ROI masks. This allowed ensuring stable image quality even with local exposure variations.

Quantitative experiments on coordinated frame sets showed stable improvement in classical quality metrics. The average PSNR increase for fused images was 3.8–4.5 dB, clearly demonstrating an improvement in signal-to-noise ratio. The structural similarity index SSIM increased on average by 15–20%, confirming that the resulting frames have higher similarity to established references. At the same time, MSE/RMSE errors decreased by 12–14%, and mean absolute error (MAE) was reduced by approximately 10%. Additional information analysis showed an increase in Shannon entropy and an increase in mutual information, indicating a larger volume of useful data in the final image.

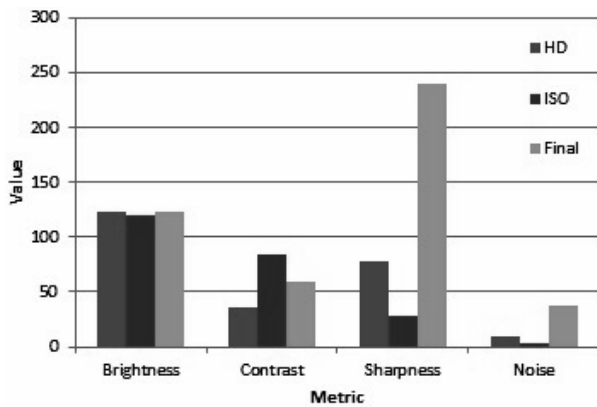
Particularly indicative was the comparison of specific fusion methods, the results of which are presented in Table. Simple weighting, wavelet-fusion, and Laplacian pyramid demonstrated relatively low PSNR values (around 17–19 dB) and did not provide sufficient detail preservation. In contrast, the proposed exposure-fusion showed the best balance of metrics: PSNR $\approx$ 29.2 dB,

SSIM = 0.78, average gradient sharpness ≈ 85.9 . Thus, it not only exceeded alternative methods but also preserved natural exposure at the level of the HD reference.

Comparison of results of different image fusion methods by main metrics

Fusion method	PSNR (dB)	SSIM (-)	MSE (-)	Sharpness gradient (-)
Simple weighting	19.7	0.61	0.015	57.7
Wavelet-fusion	19.7	0.63	0.014	57.7
Laplacian pyramid	17.3	0.58	0.017	50.6
Exposure-fusion (proposed)	29.2	0.78	0.009	85.9

Visualization of metric improvements (Fig.) additionally confirmed the identified trends. After image fusion, the indicators of brightness, contrast, and sharpness consistently exceeded the values of each channel separately, while the noise level increased in a controlled manner and could be additionally reduced through post-processing. This ensured a balanced compromise between informativeness and processing stability.



Comparison of quality metrics for HD, ISO (Thermal), and final fused images.

Figure 1 shows that the fused result preserves brightness at the level of the HD reference, has higher sharpness and contrast compared to the original channels, and demonstrates an improved signal-to-noise ratio. This confirms the effectiveness of the applied pipeline, where denoising, detail enhancement, and adaptive channel fusion are combined, collectively producing a balanced image with increased informativeness.

Practical tests on model video sequences under evening conditions confirmed the complementarity of the channels. The visible channel, limited by low contrast and noise, formed only partial contours and textures, while the thermal channel clearly reproduced silhouettes without fine structure. Their integration allowed recovering the

textural component and increasing edge sharpness, which ultimately increased target detection accuracy by more than 25% compared to using individual channels. In localization and tracking tasks, an increase in the number of stable ORB, SIFT, and SURF descriptors by 18–22% was recorded, which significantly improved the reliability of terrain binding under variable lighting conditions.

The obtained results prove the effectiveness of multimodal fusion of television and thermal channels in improving image quality and robustness of computer vision algorithms under low lighting. The synthesized frame combines the texture and detail of the HD channel with the lighting invariance of the thermal sensor, ensuring substantial growth in PSNR, SSIM, sharpness, and information saturation indicators. The proposed pipeline with preliminary calibration, denoising, detail enhancement, and adaptive exposure-fusion surpassed classical methods, guaranteeing a balance between quality and processing stability. Practical tests confirmed: target detection accuracy increases by more than 25%, and the number of reliable descriptors for localization grows by 18–22%, making the approach suitable for GPS-free navigation systems, as well as applications in security and monitoring.

IV. CONCLUSION

The conducted research confirmed that the combination of television and thermal channels produces a tangible effect in tasks of improving image quality under complex lighting conditions. The application of preliminary sensor calibration, interpolation-based resolution alignment, and adaptive fusion methods ensured PSNR growth of 3.8–4.5 dB and SSIM increase averaging 15–20%. At the same time, the reduction of MSE/RMSE and MAE errors confirmed the stability and reproducibility of results, while the increase in Shannon entropy and mutual information indicates an increase in useful content in the final image. Comparative analysis of fusion methods showed a clear advantage of exposure-fusion, which allowed preserving natural exposure while simultaneously enhancing structural details. Practical experiments on video sequences proved that channel integration increases target detection accuracy by more than 25% and ensures reliable operation of localization and tracking algorithms due to an 18–22% increase in the number of stable descriptors.

Promising directions for further research include optimization of weighting coefficients in the fusion process for different scene types, use of deep learning models for dynamic parameter adaptation to specific conditions, as well as extension of the approach to other spectral channels, particularly near-infrared and ultraviolet. Additional attention should be given to noise reduction issues after fusion and integration of the obtained methods in real time for drone onboard systems. Such research opens possibilities for creating autonomous navigation systems capable of operating in complete darkness or in the presence of active interference, which is

of significant importance for security, monitoring, and specialized applications.

V. CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

VI. DECLARATION ON GENERATIVE AI

During the preparation of this work, the authors used Claude in order to spelling check and translation. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

- [1] Wang, M., Xu, Z., Xu, M., & Lin, W. (2025). Blind multimodal quality assessment of low-light images. *International Journal of Computer Vision*, 133(4), 1665-1688. DOI: <https://doi.org/10.1007/s11263-024-02239-9>
- [2] Wang, Z., Li, D., Li, G., Zhang, Z., & Jiang, R. (2024). Multimodal low-light image enhancement with depth information. *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 4976-4985). DOI: <https://doi.org/10.1145/3664647.3680741>
- [3] Wu, Y., Cui, J., Niu, K., Lu, Y., Cheng, L., Cai, S., & Xu, C. (2025). RMF $\square$ ED: Real $\square$ Time Multimodal Fusion for Enhanced Target Detection in Low $\square$ Light Environments. *IET Cyber $\square$ Systems and Robotics*, 7(1), e70011. DOI: <https://doi.org/10.1049/csy2.70011>
- [4] Li, G., Ren, G., Wang, J., Zhi, M., Yu, Z., Jiang, B., ... & Guo, Q. (2025). Multimodal fusion transformer network for multispectral pedestrian detection in low-light condition. *Scientific Reports*, 15(1), 18778. DOI: <https://doi.org/10.1038/s41598-025-03567-7>
- [5] Lei, X., Liu, L., Jia, P., Li, H., & Zhang, H. (2024). Low-Light Infrared and Visible Image Fusion with Imbalanced Thermal Radiation Distribution. *IEEE Transactions on Instrumentation and Measurement*. DOI: 10.1109/TIM.2024.3453328.
- [6] Thaker, K., Chennupati, S., Rawashdeh, N., & Rawashdeh, S. A. (2023). Multispectral deep neural network fusion method for low-light object detection. *Journal of Imaging*, 10(1), 12. DOI: <https://doi.org/10.3390/jimaging10010012>
- [7] Xu, J., Liao, M., Kathirvel, R. P., & Patel, V. M. (2024, September). Leveraging thermal modality to enhance reconstruction in low-light conditions. In *European Conference on Computer Vision* (pp. 321-339). Cham: Springer Nature Switzerland. DOI: https://doi.org/10.1007/978-3-031-72913-3_18
- [8] Sun, C., Wang, X., Fan, S., Dai, X., Wan, Y., Jiang, X., ... & Zhong, Y. (2025). NOT-156: Night Object Tracking using Low-light and Thermal Infrared: From Multi-modal Common-aperture Camera to Benchmark Datasets. *IEEE Transactions on Geoscience and Remote Sensing*. DOI: 10.1109/TGRS.2025.3553695
- [9] Wang, Z., Wu, Y., Li, D., Tan, S., & Yin, Z. (2025, April). Thermal-Aware Low-Light Image Enhancement: A Real-World Benchmark and a New Light-Weight Model. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 8, pp. 8223-8231). DOI: <https://doi.org/10.1609/aaai.v39i8.32887>
- [10] Sharma, S., Rani, S., Sharma, A., & Dogra, A. (2024, September). Enhancing Low-Light Vision through Infrared and Visible Image Fusion. In *2024 3rd International Conference for Advancement in Technology (ICONAT)* (pp. 1-7). DOI: <https://doi.org/10.1109/ICONAT61936.2024.10775080>.
- [11] Hnatushenko, V., Shedlovskaya, Y., & Shedlovsky, I. (2022, May). Processing technology of thematic identification and classification of objects in the multispectral remote sensing imagery. In *International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence"*, (pp. 407-425). DOI: https://doi.org/10.1007/978-3-031-16203-9_24
- [12] Kashtan, V., & Hnatushenko, V. V. (2020, June). A Wavelet and HSV Pansharpening Technology of High Resolution Satellite Images. In *IntelITSIS* (pp. 67-76).
- [13] Hnatushenko, V., Nikulin, S., Kashtan, V., Korobko, O. (2024) Intelligent sentinel satellite image processing technology for land cover mapping, *CEUR Workshop Proceedings*, Volume: 3790, (pp. 194-203), DOI: <https://doi.org/10.33271/nvngu/2024-5/143>
- [14] Kashtan, V., & Hnatushenko, V. (2022, May). Deep learning technology for automatic burned area extraction using satellite high spatial resolution images. In *International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence"* (pp. 664-685). DOI: https://doi.org/10.1007/978-3-031-16203-9_37



Kostiantyn Abramenzov
Postgraduate Student at the
Department of Information
Technology and Computer
Engineering.
ORCID: 0009-0001-1155-1129



Dmytro Hryshchak
Candidate of Technical Sciences,
Doctorant at the Department of
Information Technology and
Computer Engineering. Dr.
Hryshchak is an expert in
information technology and
aerospace control.
ORCID: 0000-0001-8956-8468.
Scopus Author ID: 57220062399