Security
of Infocommunication Systems
and Internet of Things

2026 Vol 4, No 1

https://doi.org/10.31861/sisiot2026.1.01003

Received 23 January 2026; revised 31 March 2026; accepted 03 June 2026; published 30 June 2026

Research on Computer Vision Methods for Human Identification in Low Visibility Conditions

Dmytro Uhryn[1,*], Yuriy Ushenko[1] and Oleh Breslavskyi[2]

[1]Computer Science Department, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine
[2]Department of Computer Systems Software, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine

*Corresponding author (E-mail: d.ugryn@chnu.edu.ua)

ABSTRACT The article discusses the pressing scientific and practical problem of dependable human detection and identification in situations of low visibility, including dense fog, thick smoke, or extremely inadequate lighting. The research offers a systematic analysis of hybrid approaches based on the complementary integration of data from both the visible and infrared spectra. It is suggested that by synergizing these modalities, it may be possible to overcome the fundamental limitations of mono-spectral sensors, supporting robust performance in environments where traditional optical or thermal cameras individually fail. The article synthesizes a conceptual architectural model for a two-stream neural network specifically for unmanned aerial vehicles based on a comprehensive analysis and systematization of current sensor fusion methods. Particular focus is being placed on the theoretical development of a pre-assessment module logic. This module can dynamically switch between processing modes based on real-time image entropy analysis to make the best use of computing power. The research identifies a suitable technological stack, which includes end-to-end restoration architectures and lightweight integration blocks, such as temperature-guided lightweight fusion, adapted for the hardware constraints of embedded systems. The analysis indicates potential benefits of this hybrid approach in reducing spectral pollution and supporting adherence to data privacy regulations, particularly the General Data Protection Regulation, compared with mono-spectral alternatives. The article concludes by outlining strategic opportunities for the advancement of autonomous search and rescue systems, with a particular emphasis on the potential of federated learning principles for secure and decentralized model updates in practical settings.

KEYWORDS computer vision, re-identification, visible-infrared spectrum, image fusion, deep learning.

I. INTRODUCTION

Against the backdrop of an increasing number of natural and man-made disasters, effective identification and monitoring of human condition in environments with poor visibility is an important task. Although computer vision systems are frequently used for surveillance, their effectiveness drastically decreases in low visibility situations brought on by smoke, fire, fog, or darkness. These factors cause the visible light spectrum to be distorted or entirely blocked, making conventional RGB cameras useless.

Such obstacles make it impossible to recognize faces, silhouettes, or behavior in a timely manner, which directly threatens human life during search and rescue operations, monitoring at industrial facilities, or for military purposes. Traditional methods for image processing, such as contrast or brightness enhancement, are frequently insufficient, unreliable, and can produce artifacts that cause misidentification. This is why there is a need to develop and review intelligent, multispectral methods capable of providing stable and accurate human identification regardless of environmental conditions.

The purpose of this article is to formulate an approach for human identification in conditions of critically low visibility using modern computer vision methods. This approach should combine promising technological solutions for application in the fields of security and rescue operations.

This work is an analytical review; its original contribution is limited to a refined taxonomy of multispectral fusion, a conceptual two-stream architecture for UAVs, and their theoretical justification, none of which is claimed to be experimentally validated here.

II. LITERATURE REVIEW

To ensure the relevance and objectivity of the review, sources were selected from leading scientometric databases, including IEEE Xplore Digital Library and Scopus, which index key conferences on computer vision CVPR and ICCV and specialized journals TPAMI and Neurocomputing. The time frame of the review covers the period from 2015 to 2025, with a priority on research from the last 5 years, which allows us to trace the evolution of methods from classical approaches to modern deep architectures. The search was performed using combinations of keywords such as "computer vision", "re-identification", "visible-infrared spectrum", "image fusion", "deep learning", and "low visibility." Sources were included if they were peer-reviewed and directly addressed detection, re-identification, image enhancement, or multimodal fusion under degraded-visibility or hardware-constrained conditions. The selected methods were then compared along unified criteria - category and fusion level, principal advantages, inherent restrictions, and reported quantitative indicators - which form the basis of the comparative analysis. It should be noted that these quantitative indicators originate from the respective source publications and were obtained on different datasets and benchmarks; therefore, the comparison is intended to reflect general tendencies rather than a controlled head-to-head evaluation. The main focus of the literature search was on works that offer the latest methods of multispectral image fusion or cross-modal identification architectures. At the same time, in order to ensure the practical significance of the results, studies that did not contain a verified comparative analysis on publicly available datasets were not included in the analysis. Articles devoted exclusively to traditional signal processing methods without the use of deep learning technologies were also excluded, as such approaches do not provide sufficient accuracy and robustness in conditions of complex meteorological phenomena.

Analysis of the selected sources shows that the scientific community pays considerable attention to methods of primary video signal processing with moderate degradation due to fog or low light. In the field of image restoration, there has been a shift from physical scattering models to the use of deep neural networks capable of evaluating light transmission maps to restore contour sharpness [1]. At the same time, methods for improving low-light conditions are also being developed. These technologies have evolved from classic histogram equalization and retinex theory to complex CNN architectures such as RetinexNet and Zero-DCE [2]. These methods allow for the restoration of detailed facial features while minimizing information loss, but their effectiveness is critically reduced in the presence of thick smoke or complete darkness.

Another challenge is the lack of annotated data for complex conditions. Traditional teacher-led training on synthetic data often leads to overfitting. A viable approach is self-supervised learning and learning without paired data. In particular, for Re-ID tasks, the use of contrastive learning is becoming relevant, which allows the model to learn invariant features by maximizing the similarity between different augmentations of a single image without explicit class labels. Generative networks that learn in an uncontrolled mode are also used, adapting to real domains without strict binding to reference images [3].

Researchers justify the need to use the infrared spectrum for working in conditions of extremely low visibility. It has been experimentally confirmed that long-wave IR radiation (9.3 μm) effectively passes through aerosol obstacles, allowing the silhouette of a person to be distinguished [4]. However, the monospectral approach has a significant drawback – the loss of texture data. In [5], it is emphasized that the solution to this problem is multimodal fusion, which involves the stages of alignment and creation of a single informative representation. Modern algorithms, in particular TGLFusion [6], offer mechanisms controlled by object temperature for adaptive weight distribution between the visible and infrared spectra. This allows combining the textural information of RGB cameras with the stability of thermal imaging sensors.

A separate area of research is the development of re-identification systems based on deep convolutional neural networks. The problem of identifying a person across different modalities is complicated by a significant gap between modalities [7, 8]. Since manual annotation of such heterogeneous data is extremely resource-intensive, modern approaches increasingly integrate domain adaptation methods. In this context, generative adversarial networks and “modality confusion” strategies [9] are used not only to create intermediate representations [7], but also to synthesize complex weather conditions, which minimizes discrepancies in feature distribution without a teacher [9]. In addition to technical barriers, the implementation of UAV-based systems requires compliance with GDPR. Centralized video data collection creates risks of model inversion attacks. To resolve the dilemma between efficiency and privacy, it is advisable to use decentralized learning methods. The implementation of isolation protocols, such as “camera independence,” allows models to be trained directly on board without transmitting “raw” images. At the same time, instead of simple blurring, which destroys features, promising algorithms are “identity shift” algorithms that transform images by changing the absolute identity of a person but preserving the semantic connections necessary for a neural network [10, 11]. Despite significant progress, most existing systems implement fixed processing scenarios: either image enhancement or switching to a thermal imager. The lack of a flexible mechanism for adapting to dynamic changes in the environment limits their application in search and rescue operations.

III. PHYSICAL LIMITATIONS OF SENSORS AND MODELING OF VISIBILITY CONDITIONS

To justify an effective approach to identification in low-visibility conditions, the primary task is to analyze the physical limitations of RGB sensors and the advantages of thermal imaging systems. Traditional RGB cameras work by perceiving reflected visible light. In fog, smoke, or haze, image quality is significantly degraded due to the absorption and scattering of light by suspended particles in the atmosphere. The theoretical basis for describing this process is the atmospheric scattering model, which describes the observed intensity *I(x)* as a mixture of radiation from an object attenuated by the environment and atmospheric scattered light. The model formula looks like Eq. (1).

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

where *J(x)* is the image without fog, *A* is global atmospheric light, and *t(x)* is a transmission map that depends on scene depth and scattering coefficient. In foggy or smoky conditions, the value of *t(x)* decreases, leading to a loss of contrast and color [1]. In addition, adverse weather conditions such as rain or snow introduce additional noise and object occlusion, blurring contours, and complicating the operation of convolutional neural networks.

In low-light conditions, RGB images suffer from Poisson noise due to the small number of photons. In addition, algorithms often make mistakes due to bright light sources such as streetlights or headlights, which draw the attention of the neural network, leading to misclassifications. On the ExDARK dataset, which contains only low-light images, traditional proposal generation methods (e.g., edge boxes) demonstrate a detection rate below 70% for all types of low light, and for very dark images, this rate is even lower due to low contrast and noise [12]. On the KAIST dataset, which contains pairs of RGB and thermal images, a standard detector (ACF) trained only on color images showed an error rate of 90.17% in night conditions. In comparison, the use of a multispectral approach reduced this rate to ~64% [13].

Unlike RGB sensors, which operate in the visible spectrum (0.4 – 0.7 μm), thermal imaging cameras (LWIR) have fundamental physical advantages for human identification. First, long-wave infrared sensors operate in the 7.5 – 13 μm range. This is critically important because the human body emits thermal energy in this range (the peak emission is at 9.3 μm). This allows thermal imagers to capture the object's own radiation rather than reflected light, making them independent of external lighting [4, 8]. In addition, due to its longer wavelength, infrared radiation is less susceptible to scattering by smoke or fog than visible light. Studies show that thermal imaging cameras are more resistant to interference from car headlights and street lights, which blind RGB cameras Also, when the temperature drops at night, the difference in temperature between the environment and a person is even bigger than during the day, as shown in Fig. 1. This allows for clear silhouettes of people to be obtained in complete darkness, whereas in RGB images, shapes become blurred [4].

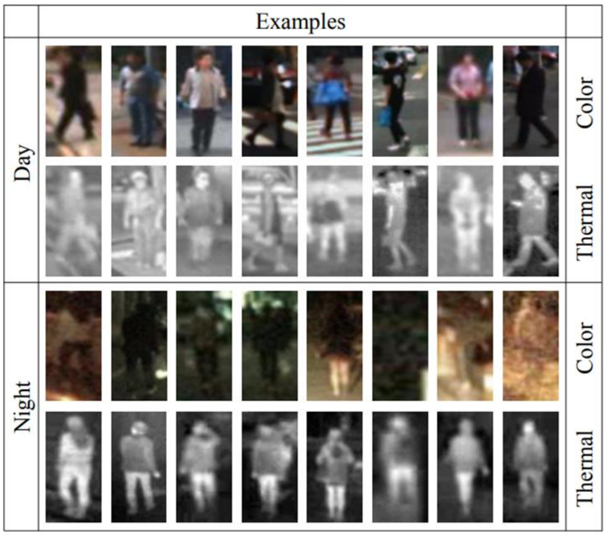

FIG. 1. Comparison of images of people taken during the day and at night with a conventional camera and a thermal imaging camera [4].

In conclusion, it can be argued that the integration of LWIR sensors allows a number of limitations of RGB sensors to be circumvented by registering a person's own thermal radiation, which eliminates the problem of low light levels and is resistant to atmospheric interference. However, it should also be noted that the resistance of infrared radiation to weather conditions has its physical limits.

Experimental studies in [14] demonstrate that at critical visibility reduction (in dense fog with meteorological visibility of 200 – 300 meters), atmospheric absorption and scattering cause the transmission coefficient to drop to almost zero, making image formation impossible even in the 10 μm range. In addition, with increasing distance, a blurring effect of the thermal signature is observed (Figs. 2 and 3), accompanied by a loss of clear intensity distribution peaks. This leads to a situation where the system can detect the presence of an object but is unable to recognize it due to the loss of contours.

FIG. 2. Thermal image of a person driving a car in thick fog with meteorological visibility of 400 meters from a distance of 300 meters [14].

FIG. 3. Thermal image of a person driving a car in thick fog with meteorological visibility of 400 meters from a distance of 400 meters [14].

However, relying entirely on the infrared spectrum is not a universal solution for identity verification tasks. Infrared images are characterized by a loss of chromatic information and low texture detail (e.g., patterns on clothing or facial features), which are key discriminative features for Re-ID algorithms [8].

IV. CLASSIFICATION OF APPROACHES TO MULTISPECTRAL FUSION

The heterogeneous nature of visible and thermal range signals creates the problem of "modality gap." This statistical discrepancy is clearly demonstrated in Fig. 4, which shows significant differences in the distribution of brightness gradients for the same scene. Direct mechanical combination of data can lead to artifacts and a decrease in the signal-to-noise ratio. This necessitates the development of not just a multisensory system, but a deeply integrated system that is capable of intelligently aligning streams and adaptively compensating for the weaknesses of one spectrum with the strengths of another.

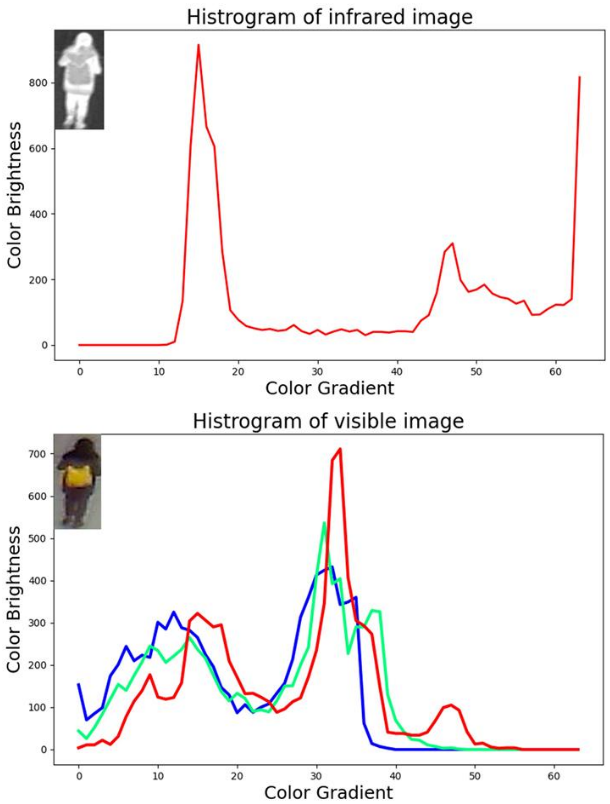

FIG. 4. Comparison of three-channel histograms of the brightness gradient of the visible spectrum for pairs of images in the visible and infrared ranges, illustrating the complexity of combining these data [8].

The effectiveness of such compensation directly depends on the selected level of data integration. Traditionally, three architectural levels of fusion are distinguished in the literature (Fig. 5):

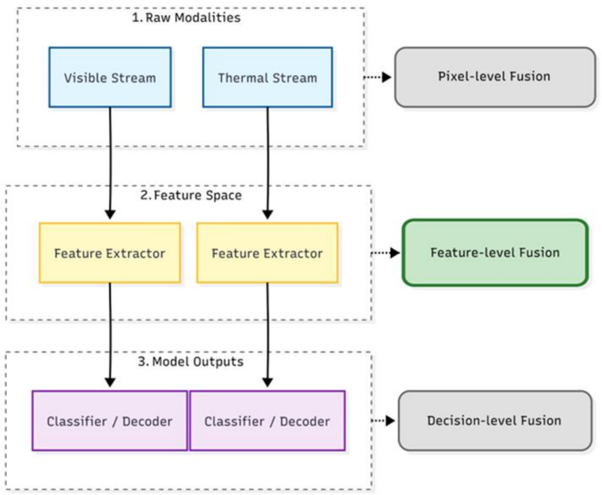

FIG. 5. Taxonomy of methods for multimodal fusion of visible and infrared spectrum images. Compiled by the authors based on [5].

- Pixel-level fusion: combines "raw" images at the network input. This method is the simplest, but it is vulnerable to spatial shifts and noise that can create visual artifacts.

- Decision-level fusion: aggregates the results of independent classifiers. At the same time, this method often ignores important correlations between modalities in the early stages, losing complementary context.

- Feature-level fusion: considered the "golden mean" and involves independent processing of each stream by its own encoder to obtain high-level vectors before their integration. Although this approach requires significant computational resources, it is the most promising strategy for Re-ID tasks [5].

While these architectural levels describe where fusion takes place, classifying methods by this dimension alone is insufficient. We therefore propose extending the taxonomy with a second, orthogonal axis – the domain-adaptation strategy – which resolves the problem of differing feature distributions and, to the best of our knowledge, is not jointly considered in prior classifications. A critical direction of classification is the domain adaptation strategy, which solves the problem of different feature distributions. The traditional approach to image generation (image-to-image translation via GAN) has significant drawbacks: high computational cost and the risk of introducing generative artifacts. A more reliable modern strategy is feature-level adaptation. Instead of attempting to visually convert infrared images into visible ones, methods of training with "modality confusion" are used. This approach entails training the feature extractor in such a way as to minimize the discrepancy between distributions, making the representation modality-invariant. This allows one to focus on the semantic content of the silhouette rather than visual similarity, which is critical for avoiding errors in fog identification [9].

V. CONCEPTUAL MODEL OF HYBRID ARCHITECTURE FOR UNMANNED AERIAL VEHICLES

Based on a detailed analysis of the physical limitations of modern sensors, we synthesize a conceptual identification architecture that implements a flexible hybrid approach to data processing. Unlike the reviewed systems, which rely on fixed processing pipelines, our contribution here is the architecture itself together with the pre-assessment switching logic: the individual building blocks (e.g., AOD-Net, Zero-DCE, TGLFusion, AGW, MCLNet) are adopted from the cited works, whereas their selection, integration, and the adaptive control logic are proposed in this study. The design builds on the principle of a two-stream network, which is often referred to as "two-leg" or "two-tower." Today, this structure is recognized by the scientific community as the most effective for heterogeneous data processing tasks, as it allows each video stream, whether visible or infrared, to have its own specialized encoder. This is critically important for the pre-processing of specific modalities before their final combination, allowing the unique characteristics of each spectrum to be taken into account [5].

The central element that is intended to support the viability of the system in a dynamically changing environment is the preliminary assessment module, which implements adaptive control logic. This module operates in real time, continuously analyzing the quality of the input video signal to choose the selection of the optimal processing branch. The system dynamically switches between the less resource-intensive "RGB enhancement only" mode and the "multimodal fusion" mode, which is used in situations with critically low visibility.

The decision-making process is based on a comprehensive analysis of a number of metrics. In particular, the first step is to evaluate the information entropy, which measures the detail saturation of the image. Should the RGB frame's entropy fail to reach the necessary threshold despite prior processing, the system interprets this as an irreversible loss of texture features, which serves as an automatic trigger for activating the thermal channel. In parallel with this, a transmission map ($t(x)$) and a reference-free NIQE (natural image quality evaluator) metric are used to assess the degree of atmospheric scattering. A high NIQE value directly indicates significant distortion of the visual signal, requiring a transition to multispectral analysis [2]. An additional weighting factor is the proportion of high temperatures (r), which calculates the fraction of pixels in the infrared image that exceed a specified temperature threshold. This value plays a key role in data fusion, as it determines the weighting coefficients (W_{ir}, W_{vi}): the higher the value of r, the greater the priority the system gives to the infrared channel, effectively compensating for the degradation of the optical spectrum [6].

The practical algorithmic implementation of the proposed architecture relies on the use of a cascade of specialized deep neural networks adapted for three key processing stages. A detailed structural-logical diagram of the algorithm is shown in Fig. 6.

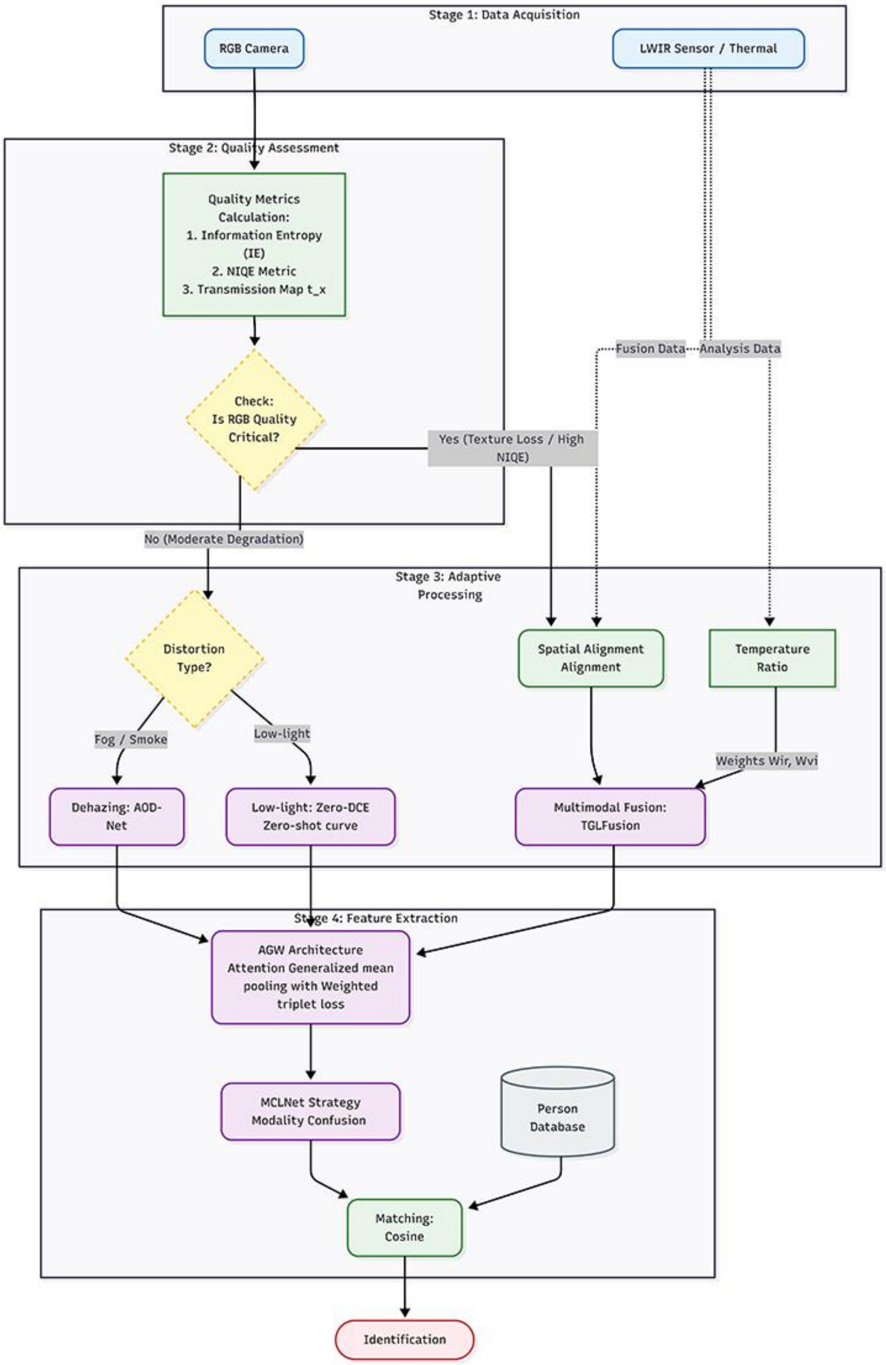

FIG. 6. Structural-logical diagram of the proposed conceptual approach to hybrid data processing. Developed by the authors.

At the image-enhancement stage, activated under moderate degradation, the approach depends on the interference type. For fog removal, end-to-end architectures such as AOD-Net have a clear advantage: unlike earlier models that estimated atmospheric parameters separately and accumulated errors, it generates parameters for a combined K-estimation, reformulating the physical scattering model for more accurate recovery [1]. In low-light conditions, Zero-DCE is promising – based on self-supervised learning, it formulates the task as estimating a pixel-correction curve via non-referential loss functions (Lspatial, Lexp), enabling training without paired datasets [2]. Both have limits, however: physical-model algorithms like AOD-Net often produce dull results due to the simplistic scattering model, while Zero-DCE struggles to preserve naturalness, with processed images prone to overexposure and an unnatural bluish tint, confirmed by low structural-similarity metrics [3].

Under poorer visibility, the system switches to multimodal fusion. Since cameras on a drone's suspension often differ in field of view, the primary task is spatial alignment via attention mechanisms, after which fusion occurs at the more efficient feature level. Here TGLFusion is most relevant for resource-limited UAVs, as its MICM module integrates thermal contours into RGB images. Contrary to the view that classical algorithms are being displaced, they remain critical in embedded systems: histogram equalization or HOG-based detectors serve as highly efficient preprocessors [15] thanks to low computational complexity. In FPGA tracking, optimization is often achieved through bottom-up design, integrating classical approaches directly into the deep-network pipeline as regularizers [16]. At the hardware level, this is typically organized around reusable modules (bundles) realized as a single customized IP block shared across layers to economize on-chip resources, embedded via a multi-stage architecture with ping-pong buffering that overlaps data-transfer latency with computation [16]. TGLFusion thus implements "lightweight" deep learning: MobileBlock blocks (like the SkyNet architecture [16]) use separable convolutions, cutting computational load while maintaining on-board accuracy.

The final stage is cross-modal identification, aimed at leveling the "modality gap." An effective solution is the AGW (Attention Generalized mean pooling with weighted triplet loss) architecture, combining non-local attention and generalized pooling to capture both global and fine-grained details. To further improve accuracy, the MCLNet (Modality confusion learning) strategy minimizes distribution discrepancies. The resulting vector is compared against the wanted-persons database using cosine-distance metrics, supporting reliable identification even in difficult conditions [9, 17]. Such mechanisms involve compromises: strict convergence to a single center (as in Center Loss) risks losing unique details needed to distinguish similar objects, so marginal restrictions must be introduced (as in the ICA method). Even powerful architectures like AGW rely primarily on global features, limiting their effectiveness under significant occlusions, where local details are critical [18].

VI. THEORETICAL JUSTIFICATION OF THE METHOD

The architectural model and algorithmic implementation proposed above rest on the concept of complementarity – the mutual reinforcement of visible and infrared information flows – a key factor in building reliable computer vision systems. The justification that follows is analytical: we reason from the physical and information-theoretic properties reported in the cited studies rather than from our own experiments. Since visible and infrared sensors respond to different physical phenomena, correctly combining the two streams creates a synergy effect. From the standpoint of image-recognition theory and machine learning, integrating two heterogeneous streams increases robustness to unpredictable changes in the environment – particularly important under partial occlusion and long-distance monitoring, where visual information becomes blurred and loses high-frequency detail.

Empirical data from comparative experiments with multispectral detectors (e.g., feature-channel combinations such as ACF+T+THOG) demonstrate a stable reduction in the average error rate of 15% or more compared to classic single-channel RGB detectors (Fig. 7). This combination synergistically fuses heterogeneous physical characteristics of an object into a single feature vector, where:

- ACF (Aggregate Channel Features) – describes the visible appearance, integrating color-space channels and gradient magnitudes to capture texture boundaries.

- T (Thermal intensity) – records absolute energy-emission values, localizing the object by its temperature contrast.

- THOG (Thermal HOG) – analyzes the geometric structure of thermal spots, highlighting specific gradients of the human form in the infrared spectrum.

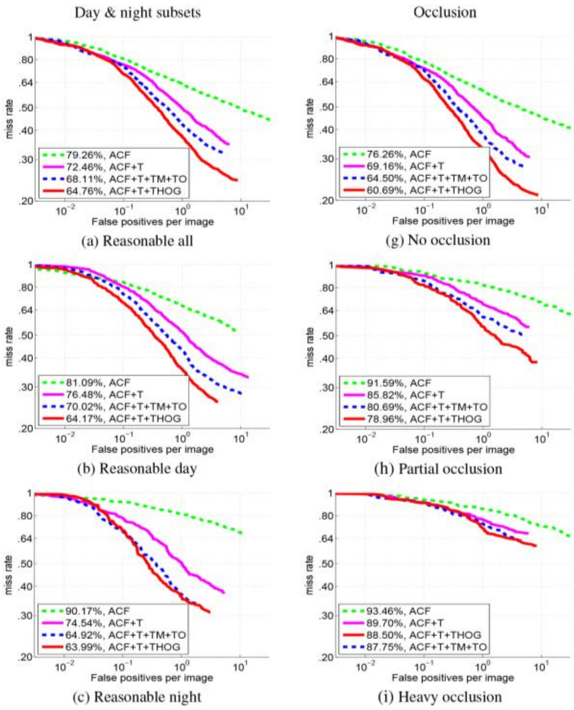

FIG. 7. Quantitative assessment of the effectiveness of multispectral detectors compared to the basic RGB method (ACF). The graphs illustrate the dependence of the error level on lighting conditions and the degree of object occlusion [4].

Such deep integration allows the algorithm to construct a multidimensional feature space where the geometric contours of the body are verified by its thermal signature, making the detector resistant to visual camouflage or complex backgrounds [4]. Because the proposed method relies on the geometric parameters of the silhouette and thermal signature rather than facial biometrics, it is designed to align with Privacy-preserving Re-ID requirements, aiming to allow use in the civil sector in compliance with GDPR, as identification occurs without processing sensitive personal data. This architecture is also "Federated-ready": the model can be trained on local nodes without transferring the video stream to the server. To protect local weight updates on board, the design relies on the "camera independence" domain-isolation principle and a secure aggregation (SecAgg) protocol, which combines the local gradients of individual nodes without exposing any single device's raw updates. Differential privacy is applied, adding calibrated noise that limits the risk of reconstructing personal data through model-inversion attacks.

Theoretically, this resistance to interference is explained by the natural filtering of background noise that occurs when spectra overlap. A detailed analysis of the classifiers' so-called "voting maps" reveals a telling pattern: models trained exclusively on visual data often mistakenly rely on features outside the human silhouette, responding to complex, textured backgrounds. Introducing an IR channel focuses the neural network's attention directly on the human body, especially its upper part, where thermal radiation is most intense. In effect, the thermal channel acts as a high-frequency spatial filter that removes information noise and amplifies the useful signal, allowing the detector to ignore visual artifacts [4].

Optimizing the signal-to-noise ratio during data fusion also deserves special attention, being non-trivial due to the different nature of the signals. Simple superimposition or mechanical averaging of images, as noted in the description of the TGLFusion method, carries a high risk of "spectral contamination." This occurs when noise artifacts from one spectrum – for example, severe blurring of an RGB frame in twilight or fog – are superimposed on a high-quality image from another, degrading the result and "polluting" the feature vector. That is why the hybrid approach described above is based on the principle of complementary restoration: clear thermal contours are used not merely as an additional layer of information, but as a structural framework for restoring or amplifying a weak RGB signal [6].

However, the problem is aggravated by the physical context of the carrier platform. UAV flight dynamics create specific distortions that cannot be simulated in laboratory conditions. The main problem is the desynchronization of modalities: the physical distance between the sensors on the drone's suspension causes a parallax effect, especially critical at short distances and not eliminable by standard calibration. High-frequency rotor vibrations add nonlinear image blurring, creating "ghost" contours at object edges that can disorient the neural network's attention mechanisms [18]. Energy consumption is another limiting factor: powerful graphics accelerators are unacceptable for drones due to limited battery life (efficient FPGA networks consume ~7 – 8 W versus >13 W on embedded GPUs). These values are based on practical power measurements reported for the SkyNet architecture [16], obtained on a Xilinx Zynq UltraScale+ (Ultra96) FPGA (≈7.26 W) and an NVIDIA Jetson TX2 embedded GPU (13.50 W) within the low-power DAC-SDC benchmark. Bandwidth limitations also make high-resolution video streaming unreliable, so most computations – in particular feature detection and extraction – must be performed on board.

Thus, the key theoretical challenge that the proposed logic aims to address is the "information imbalance" inherent in mobile systems. In real-world drone operation, channel informativeness is dynamic and stochastic: one modality can instantly become a source of pure noise, and static fusion methods would then lead to fatal blurring of the target object. Modern attention mechanisms and dynamic weight distribution based on temperature-ratio analysis (r) allow the system to adapt to input data in real time, deciding how much to "trust" each sensor at a given moment. This is intended to enable selective integration at the level of deep features, where priority is automatically given to the channel with the most discriminative information.

Such an adaptive mechanism is expected to improve identification quality compared to simple averaging methods. A generalized comparative analysis of the architectures considered, their main advantages, and quantitative indicators is presented in Table 1.

TABLE 1. Comparative analysis of computer vision methods and architectures for identification tasks in conditions of limited visibility, by category / fusion level, advantages, restrictions, and reported quantitative indicators.

Method name	Category / Level	Advantages	Restrictions	Quantitative indicators
AOD-Net [1]	Dehazing (Pixel-level)	Direct image generation via *K*-estimation, avoiding cumulative errors.	Dependence on paired data; ASM models may produce dull results.	Evaluated using PSNR/SSIM metrics (comparative data not available)
Zero-DCE [3]	Low-light enhancement (Pixel-level)	Zero-shot learning; ultra-high processing speed.	Risk of overexposure, bluish tint, noise introduction.	RunTime: 0.003 c; Params: 0.079 M; FLOPs: 84.99 G
TGLFusion [6]	Multispectral Fusion (Feature-level)	Weight distribution by temperature; ultra-light architecture (MobileBlock) for embedded systems.	Dependence on the temperature threshold (*T*) during segmentation.	Model Size: 0.229 MB; FLOPs: 34.5 G; Params: 57k
FusionGAN [6]	Multispectral Fusion (Generative)	Effectively highlights thermal targets in high-temperature scenes.	Heavy model; loss of details due to a single discriminator; blurring.	Model Size: 3.698 MB; FLOPs: 551.0 G; Params: 924k
AGW [17]	Re-ID Baseline (Feature-level)	Effective capture of small details (GeM pooling); versatility.	Substandard in partial identification (Partial Re-ID) and occlusions.	Rank-1: 70.05% (RegDB Visible-Thermal); mAP: 87.8% (Market-1501)
MCLNet [9]	Cross-Modality Re-ID (Feature-level)	Minimizing the modal gap without GAN; preserving feature diversity (ICA).	Uses mainly global features; lower efficiency than local methods.	Rank-1: 65.40% (SYSU-MM01); Rank-1: 80.31% (RegDB)

VII. CONCLUSION

The paper provides a systematic review and analysis of methods for human identification in conditions of limited visibility, resulting in the formulation of a concept for a hybrid computer vision system for unmanned platforms. The scientific novelty of the research lies in the systematization of existing approaches through the prism of specific limitations of onboard systems, which made it possible to propose a refined taxonomy of multispectral fusion methods. In contrast to traditional classifications, it takes into account the criteria of computational efficiency and vibration resistance. Based on a critical comparative analysis, it is suggested that for real-time tasks, a viable strategy is to abandon resource-intensive generative models in favor of adaptive fusion at the feature level, controlled by temperature indicators. The practical value of the work lies in the definition of a technology stack that integrates lightweight architectures and methods of domain adaptation, which enables the implementation of a search system on embedded devices. This creates a theoretical basis for the development of autonomous search and rescue systems intended to reduce spectral contamination and to support reliable identification while aiming to comply with privacy requirements.

Based on the identified trends, future research should resolve key challenges in multispectral identification through the following directions:

1. Hardware synchronization and stabilization: solving the problem of modality desynchronization in dynamics, specifically mitigating the parallax effect between RGB and IR sensors and compensating for image degradation caused by HF rotor vibrations on mobile platforms.
2. Algorithmic robustness in extreme weather: overcoming the physical limits of infrared transparency in dense fog or rain that create "blind spots," and generating real-world paired datasets to enhance model generalization beyond synthetic data limitations.
3. Privacy-centric identification metrics: transitioning from facial recognition to the analysis of anonymized patterns, such as silhouette, gait, and thermal signatures, to comply with GDPR regulations and developing corresponding effectiveness evaluation metrics.
4. Deployment of adaptive edge systems: redirecting attention over the next 3 – 5 years towards federated learning, allowing UAVs to process data locally without transmitting video streams to a server, ensuring high autonomy and data privacy.

AUTHOR CONTRIBUTIONS

D.U. – conceptualization, validation, writing-review, and editing; Y.U. – methodology, supervision; D.U., O.B. – investigation, formal analysis, visualization, writing-original draft preparation.

COMPETING INTERESTS

The authors have no competing interests.

REFERENCES

[1] J. Gui *et al.*, "A comprehensive survey and taxonomy on single image dehazing based on deep learning," *ACM Comput. Surv.*, vol. 55, no. 13, Art. no. 279, pp. 1–37, 2023.
[2] H. Tang *et al.*, "Low-illumination image enhancement based on deep learning techniques: A brief review," *Photonics*, vol. 10, no. 2, p. 198, 2023.
[3] J. Guo *et al.*, "A survey on image enhancement for low-light images," *Heliyon*, vol. 9, no. 4, p. e14558, 2023.
[4] S. Hwang *et al.*, "Multispectral pedestrian detection: Benchmark dataset and baseline," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 481–488.
[5] S. Li *et al.*, "Multimodal alignment and fusion: A survey," *arXiv preprint arXiv:2411.17040*, 2024.
[6] B. Yan *et al.*, "TGLFusion: A temperature-guided lightweight fusion method for infrared and visible images," *Sensors*, vol. 24, no. 6, p. 1735, 2024.
[7] D. Li *et al.*, "Infrared-visible cross-modal person re-identification with an X modality," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 11360–11367.
[8] H. Zheng *et al.*, "Visible-infrared person re-identification: A comprehensive survey and a new setting," *Electronics*, vol. 11, no. 3, p. 454, 2022.
[9] X. Hao *et al.*, "Cross-modality person re-identification via modality confusion and center aggregation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 16403–16412.
[10] Y. Jiang *et al.*, "From laboratory to real world: A new benchmark towards privacy-preserved visible-infrared person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.
[11] S. Dou *et al.*, "Towards privacy-preserving person re-identification via person identify shift," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 10, pp. 5821–5834, 2023.
[12] Y. P. Loh and C. S. Chan, "Getting to know low-light images with the exclusively dark (ExDark) dataset," *Computer Vision and Image Understanding (CVIU)*, vol. 178, pp. 30–42, 2019.
[13] Z. Jing, S. Li, and Q. Zhang, "YOLOv8-STE: Enhancing object detection performance under adverse weather conditions with deep learning," *Electronics*, vol. 13, no. 24, Art. no. 5049, 2024.
[14] A. F. A. Raheem, A. A. Raheem, and F. K. Fuliful, "Investigation of thermal imaging under bad weather conditions," *AIP Conference Proceedings*, vol. 2386, p. 070004, 2022.
[15] J. Cao *et al.*, "From handcrafted to deep features for pedestrian detection: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 44, no. 9, pp. 4913–4934, 2022.
[16] X. Zhang *et al.*, "SkyNet: A hardware-efficient method for object detection and tracking on embedded systems," in *Proceedings of the 3rd Conference on Machine Learning and Systems (MLSys)*, vol. 2, 2020, pp. 1–16.
[17] M. Ye *et al.*, "Deep learning for person re-identification: A survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
[18] S. Zhang *et al.*, "Person re-identification in aerial imagery," *IEEE Trans. Multimedia*, vol. 23, pp. 281–291, 2021.

Dmytro Uhryn

Graduated from Yuriy Fedkovych Chernivtsi National University, Chernivtsi. He is currently a Doctor of Technical Sciences, Professor at Yuriy Fedkovych Chernivtsi National University. Research interests: data mining, decision support information technologies, swarm intelligence systems, industry-specific geographic information systems.

ORCID ID: 0000-0003-4858-4511

Yuriy Ushenko

Prof., Computer Science Department, Chernivtsi National University, Chernivtsi, Ukraine. Research Interests: Data Mining and Analysis, Computer Vision and Pattern Recognition, Optics & Photonics, Biophysics.

ORCID ID: 0000-0003-1767-1882

Oleh Breslavskyi

PhD student, Department of Computer Systems Software, Chernivtsi National University, Chernivtsi, Ukraine. Research interests: computer vision, deep learning, embedded systems, and video analytics.

ORCID ID: 0009-0005-5164-2913

Дослідження методів комп'ютерного зору для ідентифікації людини в умовах низької видимості

Дмитро Угрин[1,*], Юрій Ушенко[1], Олег Бреславський[2]

[1]Кафедра комп'ютерних наук, Навчально-науковий інститут фізико-технічних та комп'ютерних наук, Чернівецький національний університет імені Юрія Федьковича, Чернівці, Україна

[2]Кафедра програмного забезпечення комп'ютерних систем, Навчально-науковий інститут фізико-технічних та комп'ютерних наук, Чернівецький національний університет імені Юрія Федьковича, Чернівці, Україна

*Автор-кореспондент (Електронна адреса: d.ugryn@chnu.edu.ua)

АНОТАЦІЯ У статті розглядається актуальна наукова та практична проблема надійного виявлення та ідентифікації людей в умовах поганої видимості, включаючи густий туман, сильний дим або вкрай недостатнє освітлення. Дослідження пропонує систематичний аналіз гібридних підходів, заснованих на взаємодоповнюючій інтеграції даних як з видимого, так і з інфрачервоного спектрів. Показано, що завдяки синергії цих методів потенційно можна подолати фундаментальні обмеження моноспектральних датчиків, що може сприяти надійній роботі в умовах, де традиційні оптичні або теплові камери окремо не справляються. У статті синтезовано концептуальну архітектурну модель двопотокової нейронної мережі, спеціально призначеної для безпілотних літальних апаратів, на основі комплексного аналізу та систематизації сучасних методів об'єднання даних датчиків. Особлива увага приділяється теоретичній розробці логіки модуля попередньої оцінки. Цей модуль може динамічно перемикатися між режимами обробки на основі аналізу ентропії зображення в режимі реального часу, щоб максимально ефективно використовувати обчислювальну потужність. Дослідження визначає доцільний технологічний стек, який включає архітектури відновлення «від кінця до кінця» та легкі інтеграційні блоки, такі як легка інтеграція з температурним керуванням, адаптовані до апаратних обмежень вбудованих систем. Аналіз вказує на потенційні переваги цього гібридного підходу в зменшенні спектрального забруднення та сприянні дотриманню вимог щодо конфіденційності даних, зокрема Загального регламенту про захист даних, на відміну від моноспектральних альтернатив. У висновку статті окреслено стратегічні можливості для вдосконалення автономних систем пошуку та рятування, з особливим акцентом на потенціалі принципів федеративного навчання для безпечного та децентралізованого оновлення моделей у практичних умовах.

КЛЮЧОВІ СЛОВА комп'ютерний зір, повторна ідентифікація, видимо-інфрачервоний спектр, злиття зображень, глибоке навчання.