DOI 10.2478/ama-2025-0086

FUSION OF DAY LIGHT AND INFRARED IMAGES: A SYSTEMATIC REVIEW OF THE STATE OF THE ART IN EO/IR GIMBAL SYSTEMS

Kamlesh VERMA*, Deepak YADAV*, Anitha Kumari SIVATHANU, Senthilnathan R**, Murali G**, Ranjith Pillai R**, Rajalakshmi TS**, Vignesh SM**, Madhumitha G**, Nandhini MURUGAN****

*IRDE, DRDO, Ministry of Defence, Government of India, India
**Department of Mechatronics Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu-603203, India

kamlesh.irde@gov.in, deepakyadav.drdo@gov.in, anithaks@srmist.edu.in, senthilr4@srmist.edu.in, muralig@srmist.edu.in, ranjithr1@srmist.edu.in, rajalakt1@srmist.edu.in, vigneshs1@srmist.edu.in, madhumig@srmist.edu.in, nandhinm@srmist.edu.in

received 13 March 2025, revised 08 November 2025, accepted 09 November 2025

Abstract: The development of a next-generation EO/IR Gimbal system is so rapid, and it is crucial to enhance the defence forces that require more reliable Intelligence, Surveillance, and Reconnaissance (ISR) capabilities. The possible outcomes of the development initiatives include improved target tracking, longer detection ranges, and higher image quality- all of which are essential in surveillance applications. To obtain precise object tracking in challenging situations, combined EO and IR camera images are used. Image fusion techniques enhance the features in these images; the fused images provide better tracking and detection capabilities in difficult-to-track scenarios. This survey offers an extensive investigation of image fusion methods. The evaluation of the fused image is also described as a crucial component, offering different ways to assess the quality of both full-resolution and reduced-resolution images. Finally, this work is concluded by going over the current constraints, issues, and problems with image fusion methods, datasets, and quality evaluation.

Keywords: EO/IR gimbal systems, Image Fusion, Image Registration, Fusion rules, Multimodal sensing, Object detection and tracking

1. INTRODUCTION

EO/IR Gimbal systems consist of a gimbal-mounted Electro-Optical/Infrared (EO/IR) camera. The camera is rotated about multiple axes by means of this pivotal support. Additionally, a video tracker that can follow moving targets is also built inside. With the aid of an intelligent gimbal system, the system can track its target with its capabilities of quick image processing, precise camera positioning, and multimodal information fusion. It raises awareness to real-time situations and surveillance in various lighting and weather conditions [1]-[3]. The EO/IR Gimbal system may present high-definition images to the operator and function in tandem, contingent upon favourable daylight circumstances. There are two categories of EO/IR systems: imaging and non-imaging. While non-imaging EO/IR systems are primarily concerned with long-range target surveillance, imaging EO/IR systems are committed to scanning the structure of targets and providing noise-free images for precise detection. The effectiveness of the EO/IR imaging system depends on contrast, luminance, noise, sampling, and blur. Images may suffer from distortion, occlusion, motion blur, or obstructed vision in bad weather (such as fog or rain). More sophisticated image processing and computer vision methods are needed to address these exploitation issues.

Electro-Optical and Infrared imaging sensors support information flow by highly advanced cameras that create streams of detailed images. These images are vital because they offer the user valuable information about the intended target [4]-[5]. The defining characteristics of EO and IR images are:

- EO cameras, commonly employed in surveillance, computer vision, and photography under well-lit environments, record reflected or emitted light within the visible spectrum. In contrast, Infrared cameras are used for temperature measurement, night vision, and heat signature identification because they capture the thermal radiation emitted by objects.
- Resolution is a performance parameter that defines how well the sensor can see spatial details (how small) in the object space. The resolution of the image from the EO camera is high compared to the IR camera.
- EO image sensors exhibit low noise levels. Images obtained from IR sensors are subject to high noise levels, such as dead pixels, lines, and fixed-pattern noise [6].
- Diffraction limits the quality of an imaging sensor. Due to the shorter wavelengths of EO systems, the diffraction blur is smaller compared to IR systems.
- Sensitivity is a performance parameter that defines how well

the sensor can discriminate small radiance changes in the object space (how dim).

– Turbulence describes the time-varying temperature inhomogeneities of the atmosphere, and it is responsible for fluctuations in temporal intensity. Both EO and IR imaging are affected by turbulence, which causes image distortion (loss of detail) and blur. Similar to the EO, the IR band is sensitive to large-amplitude, lower-frequency image shaking and the linear smear. It is less susceptible to the higher-frequency, smaller-amplitude harmonics.

Different image attributes do, however, come with drawbacks. Depending on the kind of light they catch, each sort of camera (EO/IR) has a unique set of uses and functions. By combining these two modalities, an item's thermal properties, in addition to its visual appearance, may be observed. This improves object detection and scene perception. This fusion of EO and IR bands provides spherical situational awareness, long-range precision targeting.

Fusion enhancement techniques are necessary for visual systems in order to preserve feature information for scenes with a vast area while filling in the missing data for important elements [7][8]. IR images include less color and texture information than visible light images, hence they might not perform as well in many human object tracking tasks. IR and EO camera pictures were used to overcome the constraints and achieve precise human object tracking in complicated circumstances. By merging the data from the IR and EO images, image fusion produces fused images with better tracking and detection capabilities, especially in difficult-to-track scenarios. Therefore, there is a need to improve the resolution of EO images by utilising data (such as temperature) from IR images. This allows for the high-resolution EO images to be used for the purpose of differentiating objects' temperatures during the day. This is accomplished by using image fusion algorithms to superimpose the temperature information of the target objects from the infrared image onto the original RGB color (EO) images.

The underlying motivation of this paper is to provide an exploration of recent literature and offer insights into the studies of image fusion. In addition, relevant topics such as steps in image fusion, challenges inherent to image fusion, performance evaluation, limitations, and future scope for work are covered.

2. METHODOLOGY

The systematic literature review for this study employed the PRISMA approach. It covers the evolution of image fusion techniques from early spatial and frequency domain approaches to recent AI-based fusion frameworks. It resulted in the identification of 110 research papers related to the chosen topic by definition of keywords: (“multimodal”), (“RGB-Thermal”), (“Image fusion method”). These papers were further refined by utilising exclusion criteria, such as repetition of papers, recent works, not relevant to EO/IR gimbal or image fusion methods, and unavailability of full-text access, bringing the total number of papers reviewed to 64. The PRISMA flow chart is given in Fig. 1.

The present work aims is to answer the following research questions:

– How can multimodal image fusion contribute to enhancing situational awareness, object detection, and target tracking in modern EO/IR gimbal systems?
– Which public datasets are most widely used for evaluating EO/IR fusion algorithms, and what are their characteristics?
– How has the field of image fusion evolved from traditional pixel-level methods to deep learning and AI-based frameworks?
– What recent advancements in deep learning (e.g. CNN, DenseFuse) have improved the robustness and accuracy of EO/IR image fusion?
– What are the strengths and limitations of various fusion algorithms when applied to real-time surveillance and tracking tasks?
– What future research directions can improve the accuracy, speed, and adaptability of EO/IR image fusion methods for next-generation surveillance platforms?

3. STEPS IN IMAGE FUSION

Multimodal image fusion, or the fusion of images from a daylight camera and a thermal image from an Infrared (IR) camera, permits a more comprehensive and informative representation of the captured image. The integration of visual and thermal information is essential to improve decision-making and analysis. It allows for improved detection and recognition of objects during night vision or in conditions with limited visibility due to smoke, fog, etc. Image fusion combines information from a daylight camera's image and a thermal imager into a single composite image. The composite image enables viewing both the thermal signature of objects and their appearance. Fusing these modalities can provide more texture details for subsequent object detection tasks.

The formal framework for image fusion is grouped into three broad categories, namely pixel level, feature level, and decision level, as shown in Fig. 2 [9-13]. The category derives its name from the level at which the fusion occurs. Pixel-level method for image fusion integrates the data directly from the input images for further processing. The feature level method contains the extraction of relevant features, such as edges, textures, or pixel intensities, that are combined to form the supplementary merged features. The decision level method is the highest processing level of the three levels. The source images are treated one at a time to extract all information, and then, according to specific criteria, the extracted information is fused. Feature and decision level fusion employs advanced mathematical and statistical procedures using expert knowledge and probability theory to assign class labels to pixels. Image fusion at the pixel level is simple to implement and preserves most of the original data. But its performance deteriorates if affected by noise. Feature-level image fusion offers greater robustness to noise and effective in real-time processing. However, it may result in the loss of supplementary information due to data compression. Decision-level image fusion provides higher accuracy and is effective in real-time processing. Also, it is less responsive to noise. The limitation of decision-level image fusion is that it retains only sufficient information from the source image, and data compression is highest as compared to the other two methods.

The process of EO/IR image fusion involves several sequential steps to ensure accurate integration of multimodal data. These steps include image registration, image resampling, applying the selected fusion method to the resampled images using appropriate fusion rules, and performing qualitative and quantitative evaluation of the fused results. Each of these stages is described in detail in the following subsections.

DOI 10.2478/ama-2025-0086

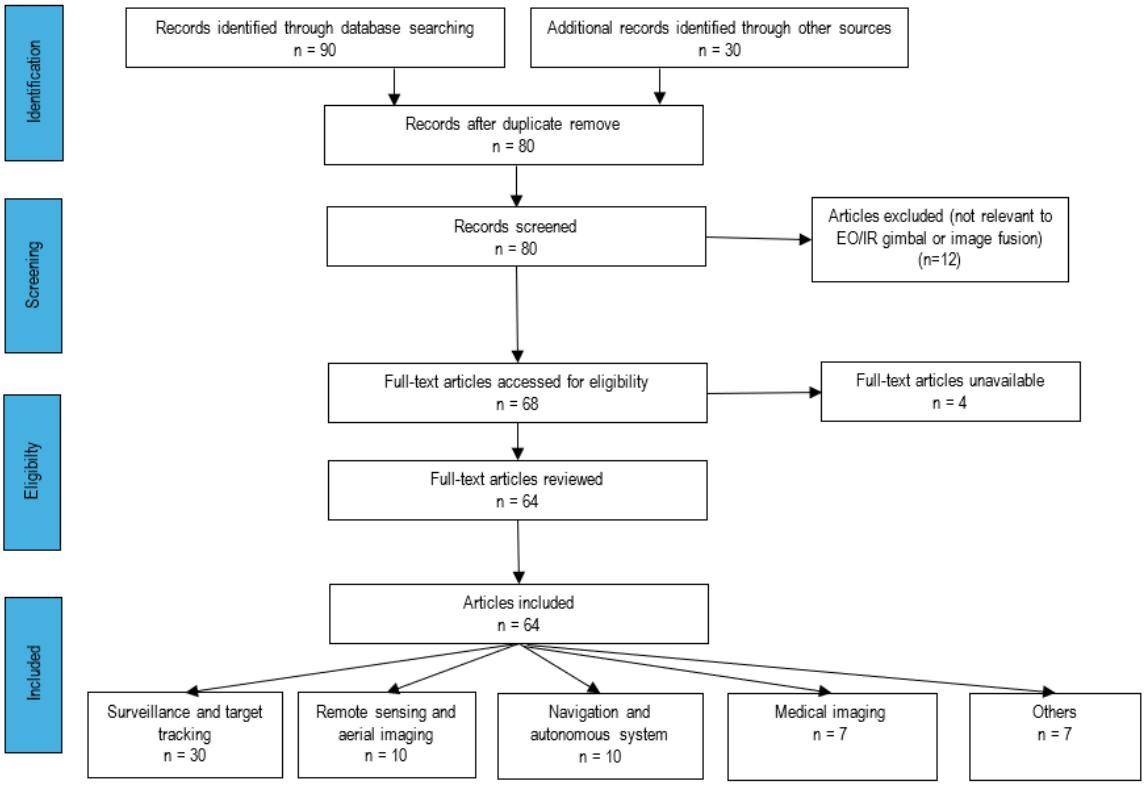

Fig. 1. PRISMA flow diagram of this systematic review

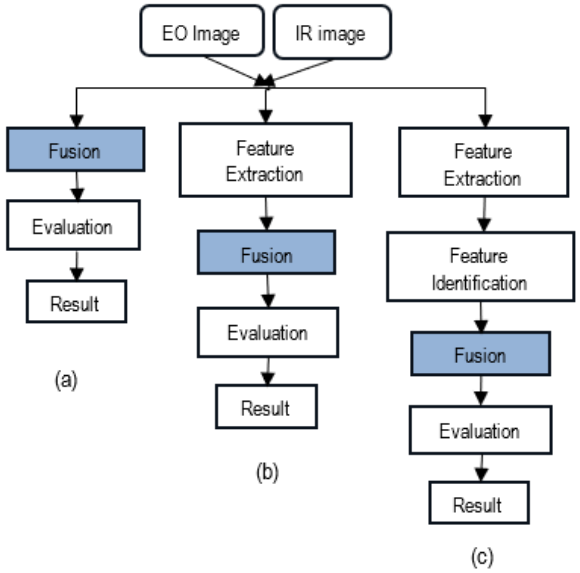

Fig. 2. Image fusion framework a) pixel level b) feature level c) decision level

3.1. Image registration

EO/IR sensors have complementary characteristics; hence, the image registration from the two different sensors is an important step. The source images should be spatially aligned to ensure that the features and objects match up accurately. Image registration can be either feature-based or image-based. As long as the image contains specific salient features, a feature-based registration approach is adopted. On the other hand, image-based registration is a more trustworthy option if the features are unreliable due to different image degradations [14][15]. This approach uses the pixel intensity without looking for the visual features.

3.2. Image resampling

Resampling is essential for images with significant resolution disparities. The pixel spacing between the EO and IR images has to be the same or a power of 2 before performing the fusion process. Conceptually, resampling involves interpolating the discrete data samples to continuous intensity, followed by sampling the interpolated image [16]-[19]. The four basic interpolation techniques used in resampling are nearest neighbor, bilinear interpolation, bicubic interpolation, and basic spline. Nearest neighbor creates a coarse block pixel with the same intensity without any new pixel formation. This technique is computationally fast but may bring significant distortion. Bilinear interpolation is a local smoothing over four neighboring pixels. It is unable to produce any undershoot or overshoot along edges. Compared to nearest neighbor interpolation, it is more complicated and requires more processing time. Bicubic interpolation increases the perceived sharpness by making pixels close to edges noticeably brighter or darker. Bicubic interpolation is generally regarded as the standard technique and is utilised in the majority of image alteration tools since it yields notably better results than the bilinear method at comparable computing costs. Cubic B-spline yields reduced smoothing of high-resolution

features in the image. When compared to bilinear interpolation algorithms, the interpolation improvement may be worth the computational load.

3.3. Image Fusion Methods

Fusion methods provide a systematic methodology for fusing images. It encompasses a wide range of techniques and strategies that are detailed in Fig. 3.

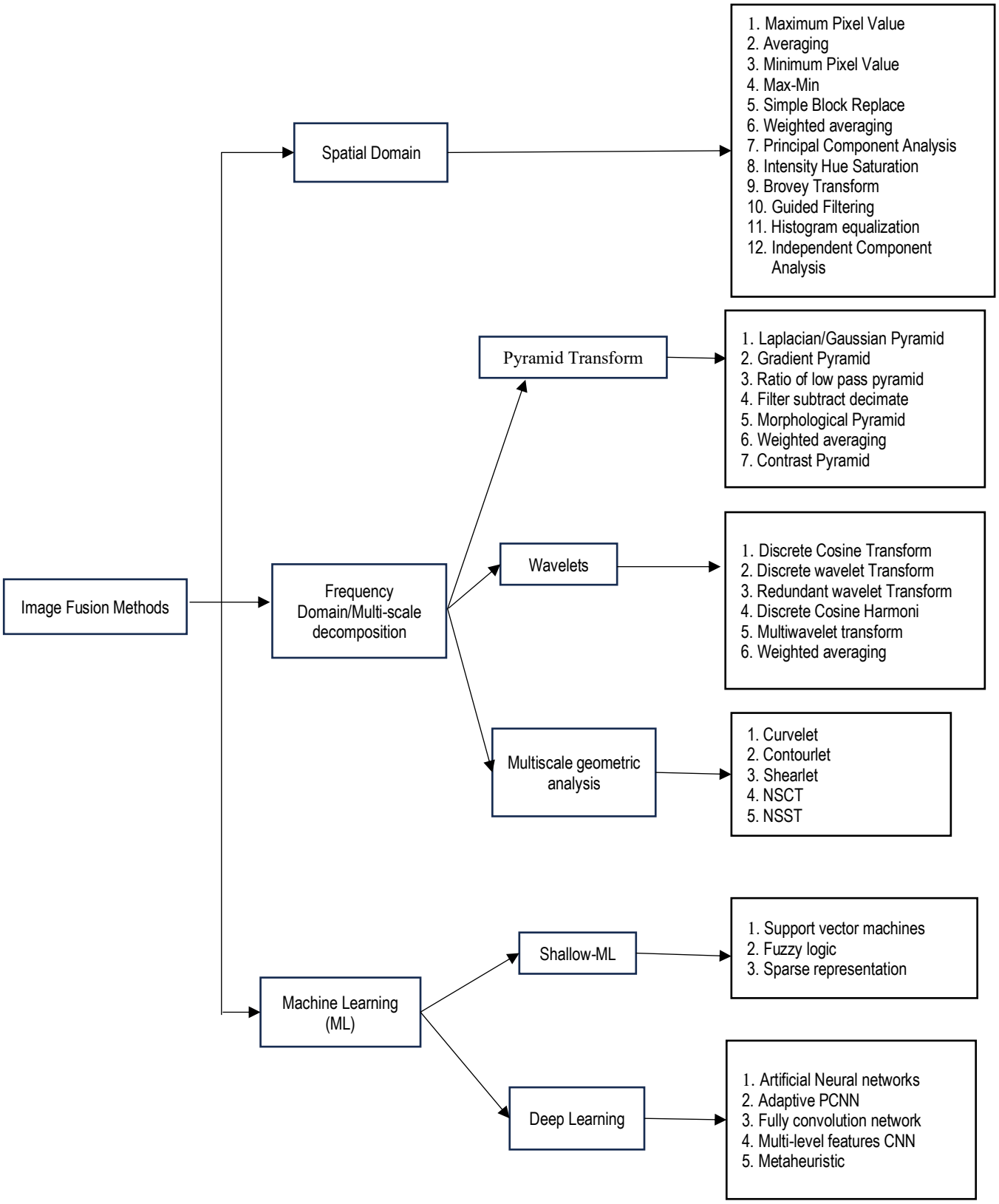

Fig. 3. Image fusion techniques

DOI 10.2478/ama-2025-0086

3.3.1. Spatial-domain image fusion methods

Spatial domain image fusion methods operate by combining pixel intensity values. These methods are simple and computationally fast. However, the quality of the fused image is not satisfactory as there are spectral deteriorations [20][21]. The merits and demerits of spatial domain fusion methods are given in Tab. 1. Techniques such as weighted averaging, Principal Component Analysis (PCA), and Intensity–Hue–Saturation (IHS) transformation have been widely used for fusing EO/IR images due to their simplicity and low computational cost. Although IHS and PCA are computationally efficient and rely on simple feature transformations, they exhibit limited adaptability to complex illumination variations. The underlying fusion rules in these methods operate primarily on intensity or component features, which can lead to spectral distortion and loss of fine detail under varying conditions.

Tab. 1. Advantages and disadvantages of Spatial Domain Methods

Fusion Method	Advantages	Disadvantages
Averaging – Image fusion by pixel averaging [22,23]	This is a basic method to identify and put into practice if the images are from the same sensor with lot of contrast and brightness. It involves a low computational cost	The fused image quality is reduced. The output images are hazy and so not suitable for real-time applications. Also, edges and image information are lost
Minimum pixel value [22]	The fused image is good if the inputs have dark shades	Fused images are characterised by low contrast and blurred
Simple block replacement [24]	Incredibly easy to understand and apply	The fused image has a random variation of brightness and color information. Fine detail of the image is less
Maximum pixel value [22,23]	The low pixel values are rejected, and the highest pixel value is used to create the fused method. So this method is susceptible to artifacts and distortion	The contrast of the fused image is decreased
Max-min [24]	Easy to implement, and the computational time is less	The efficiency of fusion is reduced, and the output image has rough edges due to blocking artifacts and isolated spots
Weighted averaging [25]	This method is easy to apply and robust.	The signal-to-noise ratio is enhanced in the fused image
	It is more suitable for multifocus images	
PCA [26,27]	This approach gives excellent spatial quality and robust	Fused images show chromatic aberration and spectral degradation
IHS [23]	The colour, resolution and features are improved in the output image. The processing time is quick with high sharpening	Only three multispectral bands are analysed. So chromatic aberration occurs in the fused image
Brovey [24]	Extremely easy and fast processing method	RGB pictures are generated with high contrast, which causes color distortion
Guided filtering [28]	This method is suitable for real-time applications and provides better performance in image smoothing	The method does not apply to sparse input data. Some edges may have halos. Also, there will be a mismatch in the color and depth details between the input and fused image

3.3.2. Transform-domain image fusion methods

Frequency domain methods are also known as Multi-Scale Decomposition (MSD) [29]-[30]. All the MSD methods involve three major steps. First, the base level features and detail level features can be analysed separately by decomposing the source image into low-frequency and high-frequency sub-bands using an appropriate multi-scale transform. The most commonly used transforms are pyramids, contourlets, discrete wavelets, shearlets and dual-tree complex wavelets. Second, the decomposed coefficients are integrated using a specific fusion rule. Third, inverse transforms are used to obtain the fused image. In the frequency domain or MSD methods, the spectral distortion is reduced and produces better SNR than the spatial domain methods. Tab. 2 shows the pros and cons in the different MSD methods. Geometric analysis-based MSD methods are effective in image representation. The most salient features in images are retained in MSD-based edge-preserving filters, including bilateral filters and guided filters. The success of transform-based methods depends on the decomposition level. If the level is low, there will be a lack of spatial details from the source image. On the contrary, if the level is high, fusion would be more sensitive to noise and it will be difficult to make accurate registration.

In the sparse transformation method, the source images are not decomposed into low-frequency and high-frequency bands, but instead, both frequency bands are assumed to have similar sparse coefficients. The sparse representations take advantage of the

regularity of source images and create coefficients of small amplitude [31][32][33]. The sparse coefficients are the important parameters that provide a final image by improving the contrast of the image by preserving the structure and visual information of the source images. However, the drawback of the sparse representation technique is that it cannot preserve fine details, and it is susceptible to misregistration errors. Despite the merits of individual fusion methods, the limitations have spurred interest in hybrid transformation strategies. Two transformation techniques, such as curvelet-wavelet, MSD-Sparse representation methods, and Principal Component Analysis - Intensity Hue Saturation, are combined in hybrid transformation methodologies.

Tab. 2. Advantages and Disadvantages of Frequency Domain Methods

Fusion Method	Advantages	Disadvantages
Morphological pyramid [34] Laplacian/Gaussian pyramid [34,35] Gradient pyramid [36] Low-pass pyramid ratio [37] Filter subtract decimate [36]	Provide better image quality	The fused image is affected by the number of breakdown levels. Also, there is no direction information, so detailed image information in different directions cannot be extracted.
Discrete cosine transform (DCT) [38]	The images are decomposed into a series of cosine waveforms representing different spatial frequency components. This compact representation makes DCT suitable for real-time applications	The fused image is blurred, and blocking artifacts are generated.
Discrete wavelet technique with Haar fusion [39]	Spectral distortions are decreased, and a fused image with better SNR is produced.	The spatial resolution of the fused image is lower. The anisotropy of the source image is not represented.
Kekre's wavelet transform fusion [40,41]	Irrespective of the size of the images, the fused image is more informative	Computation complexity is high
Kekre's hybrid wavelet-based transform fusion [42,43]	Fused image results are better with more temporal and frequency features with multi-resolution properties.	If the images are an integer power of two, this approach cannot be used
Stationary wavelet transform (SWT) [44-46]	At decomposition level 2, better results are obtained	High computational time
Curvelet Transform [47]	Best suits for edge representation	High computational time

3.3.3. AI-based and Deep Learning fusion methods

Modern technology has advanced significantly in analysis and decision-making with the incorporation of Artificial Intelligence (AI) and Deep Learning (DL) into electro-optical/infrared (EO/IR) systems [48][49]. The ability to automate and accelerate image analysis is one of the main benefits of integrating AI and DL into EO/IR systems. Large volumes of data from EO/IR sensors may be quickly processed by AI algorithms, which can then identify patterns, abnormalities, and things of interest that human operators might find difficult to identify. However, DL algorithms provide improved accuracy over time by permitting the systems to learn from data and familiarise themselves with changing environments.

Tab. 3 specifies the advantages and disadvantages of various deep learning models. The primary benefit of deep learning-based EO/IR image fusion is that it eliminates the laborious process of manually selecting parameters, demonstrates advanced performance for the complex interaction between data, and facilitates the acquisition of better fusion outcomes. In DL-based image fusion methods, Convolution Neural Networks (CNN) were widely used. The current development in the CNN-based deep learning framework [50] has been shown to be effective in handling spatial and temporal information in multimodal images. CNNs enhance the accuracy with improved computation capabilities and quantitative evaluation metrics. Also, the misregistration issues, either due to the movement of objects or the shaking of the camera, are solved. However, the efficiency is degraded in challenging situations like dark environments and bad atmospheric conditions. In dark environments, extracting the relevant features is highly challenging and in bad atmospheric conditions, such as fog, the contrast and image quality may be reduced. Furthermore, their reliance on annotated datasets restricts their suitability for real-time surveillance applications. Comparatively, Convolution Sparse Representation (CSR) offers better robustness to registration errors but remains data-intensive. Stacked Autoencoders (SAE) reduce data dependency at the cost of slower training speed and limited scalability. Overall, the critical analysis highlights a trade-off between the fused image quality, data requirement and computational efficiency, emphasizing the need for hybrid and lightweight models in challenging EO/IR environments.

Tab. 3. Advantages and Disadvantages of Deep Learning Methods

Fusion Method	Advantages	Disadvantages
Convolution Neural Network [51-53]	Features are extracted and learnt from the training data without human assistance	Computational speed is low
CSR [54]	This method is less sensitive to misregistration	Enormous training data required
SAE [55]	Limited data required for supervised learning	The model training speed depends on the processor

3.4. Fusion rules

Fusion rule is a specific guideline or mathematical procedure that dictates how information from multiple input images or sources is combined to create a single fused image [56]. A fusion rule is

DOI 10.2478/ama-2025-0086

within a fusion method to combine information from EO/IR images, and it emphasises interesting attributes while suppressing irrelevant attributes, as shown in Fig. 4. The multi-scale coefficients derived from the decomposition method are merged depending on the fusion rule. The fused image quality is highly influenced by the fusion rule. A good fusion rule leads to better results of fusion. Nevertheless, creating a single fusion rule that works for every application is not feasible.

3.4.1. Fusion rule components

Fusion rule comprises four major components: i) activity level measurement ii) coefficient grouping, iii) coefficient combination, and iv) consistency verification. The quality of each part of the input image is determined in the activity level measurement. The input images are transformed to salient features by window-based, region-based based or coefficient-based measures. In a window-based measure, a small squared window is placed over the image with the coefficient under consideration employed at the centre. Rank filter and weighted average methods are common examples of window-based measures. In coefficient-based measures, each coefficient is quantified separately. The region-based measure is parallel to the window-based measure except that region-based methods have odd shapes.

Coefficient grouping provides the details about the association between pixels of source images that are presented at the same decomposition level. The coefficient combination combines the coefficients of each image source to get the coefficients of the fused image. These rules are applied to the input image coefficient to get the final fused pixel via maximum, average or weighted average. Consistency verification ensures neighboring coefficients are fused with the same rule for a more accurate outcome.

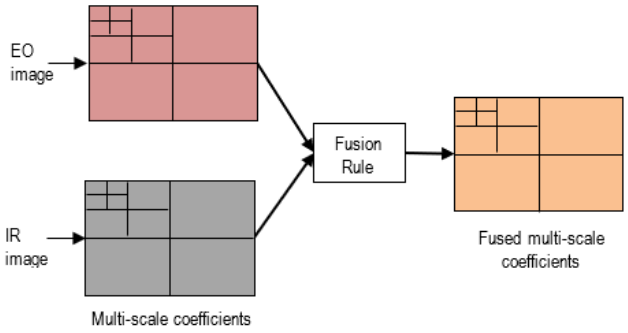

Fig. 4 Generic structure of fusion rules

4. ASSESSMENT OF PERFORMANCE AND INTRINSIC CHALLENGES IN IMAGE FUSION

Once the fusion process is complete, the performance of the fusion method adopted is assessed by two measures. The first measure is visual observation, and the second is to use evaluation metrics that involve mathematical formulas. There are different types of qualitative and quantitative evaluation metrics to evaluate the quality of the fused image. Tab. 4 lists the various evaluation metrics. The equations for the evaluation metrics can be obtained from [57,58]. Tab. 5 shows the performance evaluation metrics of quite a few fusion methods for visible and IR data sets. When paired with the aesthetic qualities of the fusion results, the congeneric values of the IY-Net algorithm appear to be acceptable even though they are not ideal. The time required for computation is a crucial factor in evaluating an algorithm's overall performance. The IY-Net algorithm is 94% faster.

Tab. 4. Performance Evaluation Metrics

S. No.	Category	Metric	Desired value for good performance	Remarks
1	Information theory	Cross entropy (CE)	Low	Evaluates the similarity of information shared between the EO/IR image and the fused image
		Entropy (EN)	High	Measures the average amount of information or detail contained in the fused image
		Mutual information (MI)	High	Quantifies the degree of statistical dependence between the source and fused images.
		Peak signal-to-noise ratio (PSNR)	High	Fused image distortion
2	Structural similarity	Universal image quality index, SSIM (Structural Similarity Index Metric)	High	Image loss (correlation loss, luminance loss) and distortion (contrast distortion)
		Root Mean Squared Error (RMSE)	Low	Calculate the variation in the source image and the fused image
3	Image feature	Average gradient (AG)	High	Insights into image clarity and fusion texture characteristics
		Edge intensity (EI)	High	Quantifies image edge intensity
		Standard deviation (SD)	High	Provide details on the factors linked with image quality - distribution of information and contrast
		Spatial frequency (SF)	High	Information on the overall activity and clarity of the image
		Gradient-based fusion performance, $Q^{AB/F}$	High	Assesses the degree to which the gradient or edge details from the source images are preserved in the fused image

Tab. 5. Quantitative results of various methods [59] – [62]

Data set	Algorithm	PSNR	SSIM	EN	MI	AG	SD	SF	Running time (s)
Multimodal image	Dense Fuse	60.27	0.72	6.84	13.67	4.24	-	-	9.85
	CNN	62.21	0.69	**7.31**	**14.67**	5.76	-	-	33.25
	ResNet	**64.23**	0.73	6.73	13.46	3.64	-	-	4.53
	Convolution Sparse representation	-	0.864	6.22	1.90	-	21.46	-	-
	Anisotropic diffusion	-	**0.94**	6.18	1.94	-	20.58	-	-
	Fourth-order partial differential equation	-	0.86	6.25	1.73	-	21.33	-	-
	Total variation and augmented Lagrangian	-	0.91	6.21	1.92	-	21.08	-	-
	Bayes Fusion	-	**0.94**	6.43	2.45	-	26.28	-	-
	Deep convolutional sparse coding	-	-	6.91	2.50	4.22	**46.97**	-	-
	DeepFuse	-	-	6.86	2.30	3.60	32.25	-	-
	Saliency Detection	-	-	6.67	1.72	3.98	28.04	-	-
	FusianGAN	-	-	6.58	2.34	2.42	29.04	-	-
	DLF	-	-	6.38	2.15	2.72	22.94	-	-
	Fast and efficient zero learning	-	-	6.63	2.23	2.55	28.09	-	-
	Discrete Wavelet Transform (DWT)	-	-	6.44	-	3.09	-	8.16	0.76
	Non-subsampled contourlet transform (NSCT)	-	-	7.17	-	5.02	-	12.78	2.03
	Multi-Focus image fusion (MFCNN)	-	-	6.61	-	3.61	-	9.55	0.38
	CNN integration (ECNN)	-	-	7.10	-	5.48	-	18.34	0.34
	Unsupervised depth model for image fusion (SESF)	-	-	**7.31**	-	**7.26**	-	**24.91**	0.31
	IY-Net	-	-	6.81	-	4.84	-	12.53	**0.16**

*Best values in Bold and 2nd best underlined

5. COMPARATIVE ANALYSIS OF THE FUSION METHODS AND INSIGHTS

A comparative assessment of traditional, multi-scale, and AI-based image fusion techniques reveals a clear evolution in both methodological complexity and fusion quality. Traditional spatial-domain approaches, such as averaging, IHS, and PCA, are computationally efficient and easy to implement but often produce fused images with blurred edges, spectral distortion, and limited robustness under varying illumination. These limitations motivated the development of multi-scale or transform-domain methods such as wavelet, contourlet, and Laplacian pyramid fusion, which provide better edge preservation and detail enhancement by separating spatial and frequency components. However, these methods still rely on manually designed fusion rules, and their performance tends to degrade when applied to dynamically changing or noisy environments.

The advent of deep learning and AI-based fusion frameworks marks a significant paradigm shift from hand-crafted feature extraction to data-driven representation learning. Convolutional neural networks (CNNs), GAN-based architectures, and hybrid deep-learning models such as DenseFuse have demonstrated substantial improvements in fusion quality, achieving higher PSNR, SSIM, and mutual information values compared to traditional and multi-scale methods. These models can automatically learn optimal fusion rules and adapt to diverse image characteristics without manual intervention. Nevertheless, their deployment in real-world surveillance and gimbal systems remains constrained by high computational demands, data dependency, and limited interpretability.

Overall, the comparative evaluation indicates a fundamental trade-off between fusion quality, computational efficiency, and interpretability. Traditional methods remain suitable for real-time or resource-limited applications, while multi-scale techniques offer a balance between performance and complexity. The progression from traditional methods to deep learning–based fusion reflects a paradigm shift from handcrafted design to data-driven optimization. Despite measurable gains in image quality, the computational burden and lack of interpretability in DNNs hinder real-time adoption. Thus, future research should focus on hybrid architectures that balance fusion accuracy, transparency, and efficiency. This analysis highlights the growing need for hybrid and lightweight fusion frameworks that can integrate the interpretability of traditional and transform-based methods with the adaptive learning capabilities of deep models. Such approaches will be critical for advancing EO/IR fusion in real-time defense, surveillance, and autonomous vision systems.

6. BENCHMARKING DATASETS

Some public databases containing visible and infrared image pairs are listed in Tab. 6. The Netherlands Organization for Applied Scientific Research (TNO) dataset comprises visual and infrared nighttime images of numerous military and surveillance circumstances. It shows different objects and targets with rural and urban backgrounds. The KAIST multispectral dataset contains color and thermal pairs taken from vehicles. All the pairs include human annotations and temporal alignment between bounding boxes. Visual and Image Fusion Benchmark (VIFB) provides a platform for a comprehensive comparison of VIF algorithms. There are 21 image pairs, 20 image fusion algorithms and 13 evaluation metrics in VIFB. Among the existing datasets, LLVIP (A Visible-infrared Paired Dataset for Low-light Vision) stands to be the largest dataset featuring 15488 spatially and temporally aligned image pairs with dark scenes.

Tab. 6. Benchmarking datasets

S.No.	Database Name	Year	Web Address
1.	TNO	2014	https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029
2.	KAIST	2015	https://soonminhwang.github.io/rgbt-ped-detection/
3.	VIFB	2020	https://github.com/xingchenzhang/VIFB
4.	LLVIP	2021	https://bupt-ai-cz.github.io/LLVIP/

7. CHALLENGES INHERENT TO EO/IR IMAGE FUSION

The challenges intrinsic to the EO/IR image fusion are

- Imperfect environmental conditions - The images might have been obtained from unfavourable conditions. So, the input images may comprise under-exposure and serious noise due to weather and illumination conditions. So, pre-processing steps such as noise reduction, radiometric calibration, and contrast enhancement [63] are required to improve the quality of the fused image.
- Object motion and misalignment - The sensors capture the images while the objects are moving. As a result, fused images are created with wraith artifacts. So, it is extremely challenging for precise and accurate image registration.
- Spectral and resolution disparities - Due to the prominent spectral difference and resolution disparities among the input images, selecting an appropriate fusion algorithm becomes crucial. The choice of algorithm significantly influences the quality and information content of the fused image.
- Computational efficiency - The image fusion algorithm must be computationally effective in merging the information from the source images to get the fused image, engaging continuous real-time monitoring.
- Target saliency preservation - Target saliency refers to the emphasis placed on specific objects or features within an image during the fusion process. The target saliency should be maintained to enhance the visibility and importance of particular elements in the fused image. The choice of fusion rules or methods plays a crucial role in achieving target saliency. The fusion rule should be chosen so that it accentuates the features or regions of interest while preserving the complementary information from both the visible and IR images.

8. LIMITATIONS AND FUTURE WORK

The major factors that are expected from the fusion methods for surveillance applications are i) The algorithms should effectively integrate the data from EO/IR images. ii) The fusion method must be computationally efficient for performing real-time surveillance. iii) The developed fusion methods should be robust to serious noise and underexposure conditions while achieving high quality fused image. Although there are various image fusion methods at present, there are still many open-ended challenges that need to be addressed.

- Dependence on image registration - The images obtained from EO/IR sensors will have some spatial errors. So it is important to perform image registration before image fusion. However, most of the image fusion algorithms are based on pre-registered images.
- Noise sensitivity - The image obtained from the IR sensor is generally disturbed by noise. Although the present time fusion algorithms have good metrics for noiseless images, their performance on noisy images has not yet been determined.
- Resolution disparity - The image resolution from EO/IR sensors is different.To achieve efficient fusion, it is difficult to get over the resolution disparity and fully utilize the information in several source images [64]. The issues related to the selection of the upsampling strategy and position of upsampling are unresolved, despite the fact that various techniques have been put forth to address varying resolution image fusion. More crucially, the goal is to naturally combine the features of image fusion and super-resolution work to create deep networks.
- Data scarcity for deep learning models - Deep learning neural networks, such as CNN, GAN, and Transformers, require large, diverse, and labelled datasets to train effectively. Thermal datasets are scarce in the constructed dataset. So, the DNN-based fusion methods may be inferior to the traditional methods due to insufficient data for training. Therefore, a larger dataset needs to be constructed for sufficient training of DNNs
- Model complexity and real-time limitations - Network architectures are becoming increasingly complex in the pursuit of improved fusion performance. However, lightweight and computationally efficient deep learning–based solutions tailored for industrial applications remain limited, posing challenges for real-time and large-scale deployments.

Future EO/IR fusion systems are expected to evolve towards:

- Transfer learning -Transfer learning by using deep neural networks has been a potential option for surveillance applications. This approach can address data insufficiency by leveraging pretrained models. It might help with the human annotation issue, which is considered to be complex and expensive. Future fusion technologies may possibly include ideas of knowledge transfer for improvement.
- Hybrid and context-aware fusion frameworks - Combining mult-scale transforms with CNNs, or integrating optimisation-based and deep learning approaches, can achieve better robustness across varying environmental conditions. Such a framework can dynamically blend fusion rules based on scene content.
- Explainable AI – Incorporating such models will enhance the interpretability of deep fusion outputs for defence and surveillance missions where transparency is critical.
- Integration with object detection and tracking pipelines – Joint fusion and perception architectures can improve object detection and tracking under low-light or occluded conditions, providing semantic understanding alongside fused imagery.
- Cross-modal attention mechanisms and transformer-based architectures – These models demonstrate the effectiveness of global information perception in multimodal image fusion. It can handle long-range dependencies and complex spatial relationships between modalities.

– Multi-sensor fusion ecosystems - Future research will likely integrate EO/IR data with radar, LiDAR, or hyperspectral imagery to achieve a holistic situational awareness.

9. CONCLUSION

The continuing challenge in surveillance operations is, maintaining a technological edge in EO/IR imaging systems. In addition to improving the individual sensor technologies, maintaining the technological edge also depends on fusing the images from the two sensors and sharing the information in real-time. Presently, it has been demonstrated that image fusion algorithms are useful tools for improving image information for visual interpretation. Pixel-level-based image fusion techniques are widely used to analyze multimodal images. In MSD fusion methods, the spatial structures are represented with wavelets, edge-preserving filtering, etc. Further, the fusion performance has been improved by considering the high correlation among neighboring pixels. Nevertheless, several issues still affect image fusion and objective fusion performance evaluation. These issues include picture noise, disparities in image resolution, unfavorable environmental circumstances, computational complexity, moving objects, and imaging hardware constraints. Thus, it is anticipated that in the years to come, new studies and useful applications based on image fusion will continue to expand.

REFERENCES

1. Artan GG, Tombul GS. The future trends of EO/IR systems for ISR platforms. InImage Sensing Technologies: Materials, Devices, Systems and Applications IX. SPIE. 2022;12091:76-88.
2. Gonzalez-Jorge H, Aldao E, Fontenla-Carrera G, Veiga-López F, Balvís E, Ríos-Otero E. Counter Drone Technology: A Review.
3. Dudek A, Stütz P. A Cloud Detection System for UAV Sense and Avoid: Flight Experiments to Analyze the Impact of Varying Environmental Conditions. InAIAA SCITECH 2024 Forum. 2024; 1859.
4. Munir A, Siddiqui AJ, Anwar S. Investigation of uav detection in images with complex backgrounds and rainy artifacts. InProceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024;221-230.
5. Ouyang Y, Shi B, Huang X, Lu L, Jiang Y. Research on infrared point target recognition method based on space-based early warning system. InInternational Conference on Algorithm, Imaging Processing, and Machine Vision (AIPMV 2023) SPIE. 2024; 12969: 697-701.
6. Kim BH, Kim MY, Chae YS. Background registration-based adaptive noise filtering of LWIR/MWIR imaging sensors for UAV applications. Sensors. 2017;18(1):60.
7. Shang X, Li G, Jiang Z, Zhang S, Ding N, Liu J. Holistic dynamic frequency transformer for image fusion and exposure correction. Information Fusion. 2024;102:102073.
8. Singh N, Bhat A. A robust model for improving the quality of underwater images using enhancement techniques. Multimed Tools Appl. 2024;83(1):2267-88.
9. Sharma M. A review: image fusion techniques and applications. Int J Comput Sci Inf Technol. 2016;7(3):1082-5.
10. Xiao X, Li C, He H, Huang J, Yu T. Rotating machinery fault diagnosis method based on multi-level fusion framework of multi-sensor information. Information Fusion. 2025;113:102621.
11. Li S, Kang X, Fang L, Hu J, Yin H. Pixel-level image fusion: A survey of the state of the art.. Information Fusion. 2017;33:100-12.
12. Liu S, Shi M, Zhu Z, Zhao J. Image fusion based on complex-shearlet domain with guided filtering. Multidimensional Systems and Signal Processing. 2017;28(1):207-24.
13. Bhutto JA, Lianfang T, Du Q, Soomro TA, Lubin Y, Tahir MF. An enhanced image fusion algorithm by combined histogram equalization and fast gray level grouping using multi-scale decomposition and gray-PCA. IEEE Access. 2020;8:157005-21.
14. Lee T, Kim S. Research trend analysis for EO-IR image registration. In: 2022 22nd International Conference on Control, Automation and Systems (ICCAS). IEEE. 2022;1288-91.
15. Velesaca HO, Bastidas G, Rouhani M, Sappa AD. Multimodal image registration techniques: a comprehensive survey. Multimedia Tools and Applications. 2024;83(23):63919-47
16. Patil MS. Interpolation techniques in image resampling. Int. J. Eng. Technol. 2018;7(3.34):567-70.
17. Han D. Comparison of commonly used image interpolation methods. InConference of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013). Atlantis Press. 2013;1556-1559).
18. Parsania PS, Virparia PV. A comparative analysis of image interpolation algorithms. International Journal of Advanced Research in Computer and Communication Engineering. 2016;5(1):29-34.
19. Kaur R, Singh S, Sethi GK. Spatial and Spectral Analysis of Resampling Algorithms in Image Fusion of Optical and Microwave Satellite Images: A Case Study Over Western Himalayas. Journal of the Indian Society of Remote Sensing. 2024;52(10):2317-34.
20. Papadopoulos S, Koukiou G, Anastassopoulos V. Decision Fusion at Pixel Level of Multi-Band Data for Land Cover Classification-A Review. Journal of Imaging. 2024;10(1):15.
21. Chen J, Chen L, Shabaz M. Image fusion algorithm at pixel level based on edge detection. Journal of Healthcare Engineering. 2021; (1):5760660.
22. Zhou M, Huang J, Yan K, Hong D, Jia X, Chanussot J, Li C. A general spatial-frequency learning framework for multimodal image fusion. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2024.
23. Mishra VK, Kumar R, Nareti U, Pant T, Soni PK. Pansharpening using IHS method on multi-sensor data and multiple feature extraction using modified Otsu thresholding. Journal of the Indian Society of Remote Sensing. 2024;52(1):113-26.
24. Jiang D, Zhuang D, Huang Y, Fu J. Survey of multispectral image fusion techniques in remote sensing applications. Image fusion and its applications. 2011;1:1-22.
25. Alhatami E, Huang M, Bhatti UA. Image fusion techniques and applications for remote sensing and medical images. InDeep Learning for Multimedia Processing Applications. CRC Press. 2024;154-175.
26. Smith LI. A tutorial on principal components analysis.
27. Li S, Kang X, Hu J. Image fusion with guided filtering. IEEE Transactions on Image processing. 2013;22(7):2864-75.
28. Zhang X, Wang X, Yan C, Sun Q. EV-fusion: A novel infrared and low-light color visible image fusion network integrating unsupervised visible image enhancement. IEEE Sensors Journal. 2024;24(4):4920-34.
29. Dogra A, Goyal B, Agrawal S. From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications. IEEE access. 2017;5:16040-67.
30. Gong X, Hou Z, Wan Y, Zhong Y, Zhang M, Lv K. Multispectral and SAR image fusion for multiscale decomposition based on least squares optimization rolling guidance filtering. IEEE Transactions on Geoscience and Remote Sensing. 2024;62:1-20.
31. Yin H, Li Y, Chai Y, Liu Z, Zhu Z. A novel sparse-representation-based multi-focus image fusion approach. Neurocomputing. 2016;216:216-29.
32. Ma X, Hu S, Liu S, Fang J, Xu S. Remote sensing image fusion based on sparse representation and guided filtering. Electronics. 2019;8(3):303.
33. Li L, Lv M, Jia Z, Ma H. Sparse representation-based multi-focus image fusion method via local energy in shearlet domain. Sensors. 2023;23(6):2888.

DOI 10.2478/ama-2025-0086

34. Li L, Shi Y, Lv M, Jia Z, Liu M, Zhao X, Zhang X, Ma H. Infrared and visible image fusion via sparse representation and guided filtering in laplacian pyramid domain. Remote Sensing. 2024;16(20):3804.
35. Singh P, Bhandari AK. Laplacian and gaussian pyramid based multiscale fusion for nighttime image enhancement. Multimedia Tools and Applications. 2025;84(15):15527-51.
36. Xi Y, Liu D, Kou R, Zhang J, Yu W. Gradient Enhanced Feature Pyramid Network for Infrared Small Target Detection. IEEE Geoscience and Remote Sensing Letters. 2025.
37. Mehta B, Patel H, Nanavati M, Limbad N, Gohel P. Implementation and comparative analysis of various Pyramid-based Image Fusion techniques for Multimodal MRI images of brain. Journal of Integrated Science and Technology. 2025;13(2):1031.
38. Naidu VP. Discrete cosine transform based image fusion techniques. Journal of Communication, Navigation and Signal Processing. 2012;1(1):35-45.
39. Singh R, Khare A. Multiscale medical image fusion in wavelet domain. The Scientific World Journal. 2013;1:521034.
40. Kekre HB, Athawale A, Sadavarti D. Algorithm to generate Kekre's Wavelet transform from Kekre's Transform. International Journal of Engineering Science and Technology. 2010;2(5):756-67.
41. Kekre HB, Sarode T, Dhannawat R. Implementation and comparison of different transform techniques using Kekre's wavelet transform for image fusion. International Journal of Computer Applications. 2012;44(10):41-8.
42. Dhannawat R, Sarode T. Kekre's hybrid wavelet transform technique with dct, walsh, hartley and kekre's transform for image fusion. International Journal of Computer Engineering and Technology (IJCET). 2013;4(1):195-202.
43. Kekre HB, Sarode T, Dhannawat R. Image fusion using Kekre's hybrid wavelet transform. In2012 International Conference on Communication, Information & Computing Technology (ICCICT). IEEE; 2012: 1-6.
44. Danyal MM, Khan S, Khan RS, Jan S, Rahman N. Enhancing Multi-Modality Medical Imaging: A Novel Approach with Laplacian Filter+ Discrete Fourier Transform Pre-Processing and Stationary Wavelet Transform Fusion. J. Intell. Med. Healthc. 2024;2:35-53.
45. Udomhunsakul S, Yamsang P, Tumthong S, Borwonwatanadelok P. Multiresolution edge fusion using SWT and SFM. InProceedings of the world congress on engineering 2011;2:6-8.
46. Naseem S, Mahmood T, Khan AR, Farooq U, Nawazish S, Alamri FS, Saba T. Image Fusion Using Wavelet Transformation and XGboost Algorithm. Computers, Materials & Continua. 2024;79(1).
47. Dong L, Yang Q, Wu H, Xiao H, Xu M. High quality multi-spectral and panchromatic image fusion technologies based on curvelet transform. Neurocomputing. 2015;159:268-74.
48. An FP, Ma XM, Bai L. Image fusion algorithm based on unsupervised deep learning-optimized sparse representation. Biomedical Signal Processing and Control. 2022;71:103140.
49. Cheng P, Xiong Z, Bao Y, Zhuang P, Zhang Y, Blasch E, Chen G. A deep learning-enhanced multi-modal sensing platform for robust human object detection and tracking in challenging environments. Electronics. 2023;12(16):3423.
50. Sreeja G, Saraniya O. Image fusion through deep convolutional neural network. InDeep learning and parallel computing environment for bioengineering systems. Academic Press. 2019; 37-52.
51. Liu Y, Chen X, Peng H, Wang Z. Multi-focus image fusion with a deep convolutional neural network. Information Fusion. 2017;36:191-207.
52. Du C, Gao S. Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network. IEEE access. 2017;5:15750-61.
53. Masi G, Cozzolino D, Verdoliva L, Scarpa G. Pansharpening by convolutional neural networks. Remote Sensing. 2016 Jul 14;8(7):594.
54. Liu Y, Chen X, Ward RK, Wang ZJ. Image fusion with convolutional sparse representation. IEEE signal processing letters. 2016;23(12):1882-6.
55. Huang W, Xiao L, Wei Z, Liu H, Tang S. A new pan-sharpening method with deep neural networks. IEEE Geoscience and Remote Sensing Letters. 2015;12(5):1037-41.
56. Xu H, Zhang H, Ma J. Classification saliency-based rule for visible and infrared image fusion. IEEE Transactions on Computational Imaging. 2021;7:824-36.
57. Singh S, Singh H, Bueno G, Deniz O, Singh S, Monga H, Hrisheekesha PN, Pedraza A. A review of image fusion: Methods, applications and performance metrics. Digital Signal Processing. 2023;137:104020.
58. Zeng Y, Wang X, Zhao H, Jin Y, Giannopoulos GA, Li Y. Image fusion methods in high-speed railway scenes: A survey. High-speed Railway. 2023;1(2):87-91.
59. Zhu P, Ouyang W, Guo Y, Zhou X. A Two-To-One Deep Learning General Framework for Image Fusion. Frontiers in bioengineering and biotechnology. 2022;10:923364.
60. Zhao Z, Xu S, Zhang C, Liu J, Zhang J. Bayesian fusion for infrared and visible images. Signal Processing. 2020;177:107734.
61. Zhang C, Hu H, Tai Y, Yun L, Zhang J. Trustworthy image fusion with deep learning for wireless applications. Wireless Communications and Mobile Computing. 2021;1:6220166.
62. Xu S, Zhao Z, Wang Y, Zhang C, Liu J, Zhang J. Deep convolutional sparse coding networks for image fusion. arXiv preprint arXiv:2005.08448. 2020.
63. Ma J, Xu H, Jiang J, Mei X, Zhang XP. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing. 2020;29:4980-95.
64. Li H, Cen Y, Liu Y, Chen X, Yu Z. Different input resolutions and arbitrary output resolution: A meta learning-based deep framework for infrared and visible image fusion. IEEE Transactions on Image Processing. 2021;30:4070-83.

This work is supported by the CARS scheme of Instruments Research and Development Establishment (IRDE), DRDO, Dehradun, India by the Ministry of Defence, Government of India under the Grant DRDO/DFMM/PL/83226/M/01/1976/D (R&D).

Kamlesh Verma: https://orcid.org/0000-0001-8821-502X

Deepak Yadav: https://orcid.org/0009-0000-1940-3932

Anitha Kumari Sivathanu: https://orcid.org/0000-0001-6185-5238

Senthilnathan R: https://orcid.org/0000-0001-5628-2279

Murali G: https://orcid.org/0000-0002-2219-1499

Ranjith Pillai R: https://orcid.org/0000-0001-9263-7954

Rajalakshmi TS: https://orcid.org/0000-0001-9090-2419

Vignesh SM: https://orcid.org/0000-0002-3501-7844

Madhumitha G: https://orcid.org/0000-0003-4811-6195

Nandhini Murugan: https://orcid.org/0000-0001-6689-2599