Journal of Computer Science

Research Article

A Multi-Modal Image Fusion Approach for Visual and Infrared Images via Shearlet-Based Decomposition

Apoorav Sharma[1], Shagun Sharma[2], Kalpna Guleria[3], Ayush Dogra[3], Pankaj Lathar[4], Archana Saini[3] and Bhawna Goyal[5]

[1]*School of Computing, Sunstone, Rayat Bahra University, Mohali, Punjab 140104, India*
[2]*School of Computing Science and Engineering, VIT Bhopal University, Sehore, Bhopal, Madhya Pradesh, India*
[3]*Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India*
[4]*Department of Computer Science and Applications, Delhi Skill and Entrepreneurship University, Delhi, India*
[5]*Department of Engineering, Marwadi University Research Centre, Marwadi University, Rajkot, Gujarat, 360003, India*

Article history
Received: 19-05-2025
Revised: 18-06-2025
Accepted: 08-07-2025

Corresponding Author:
Archana Saini
Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India
Email: saini.archana@chitkara.edu.in

Abstract: The integration of optical lens technologies with night vision has become necessary due to the increasing demand to enhance public safety and surveillance, particularly in vulnerable areas, public transportation, and airports. Due to vision clarity issues or the lack of thermal information, conventional single-modality systems often fail to detect concealed dangers. This research presents a multimodal image fusion architecture that integrates visible and infrared (IR) images to enhance hidden weapon detection, thereby mitigating these limitations. Whereas infrared images provide valuable heat signals that can penetrate clothing and reveal hidden objects depending on temperature gradients, visible photographs provide accurate spatial and textural information. In this article, an efficient VR-IR image integration model is proposed by merging distinct images acquired from different sensors: Visible Images containing high spatial are merged with infrared images containing high thermal radiation information and low spatial resolution details. The proposed fusion algorithm harnesses the attributes of the Shearlet transform (NSST) and spectral residual details information. Furthermore, the proposed architecture yields improved visual and objective results compared to other fusion algorithms. The proposed method surpasses all current methods with the highest fusion rate of 0.9276, minimum information loss (0.0536), and shows artifact (0.0128), indicating nearly no extra noise or visual distortion.

Keywords: Visual, Infrared, Fusion, Night Vision, Shearlet, Optical Lens

Introduction

Images are captured in almost every range of the Electromagnetic spectrum. The wavelength of the waves decides their level of penetration, absorption, and scattering rate. Based on this, different sensors are used to capture the information (Muñoz *et al.,* 2025). The images captured through visible light sensors capture the reflected components of visible white light hence, the visible images have clearly defined boundaries of objects and an excellent contrast if the objects in the scene are well illuminated (Zhu and Zhang, 2025). On the other hand, the Infrared sensors capture relatively larger wavelengths which can penetrate one level deeper than the visible light. The higher wavelengths help capture the nocturnal features of the scene (Hadinejad *et al.,* 2025). Since the images are captured through the radiation pattern emitted, various object boundaries are formed according to the temperature difference between different objects in the scene. This reveals the hidden object information but with low details (Khor *et al.,* 2024a).

Image fusion brings out the subtle aspects and details of an image to integrate them into a single image which may appeal to a certain set of observers for instance, it helps in better surveillance when multi-sensor images like Visible and Infrared images of the same scene are fused (Wang *et al.,* 2024). Additionally, it also improves the contrast and overall luminance of the image which

facilitates in better perception of details. As an image is simply an array of pixel intensity values, a simple averaging operation also provides a fused version, but with a lot of lost details and artifacts (Jagalingam and Hegde, 2007). The primary goal is to transfer as much useful information as possible by manipulating the source images in a controlled and efficient manner, utilising various tools such as transforms or spatial filters (Dey and Aravind, 2024). These tools are used to separate the high-frequency (edges and contours or detail) and low-frequency (consistent or base) information. The purpose of this separation can be better understood if the general framework of image fusion is known firsthand. The general framework of an image fusion technique can be described in three steps (Dong and Wang, 2024):

- Multi-scale decomposition (MSD)
- Separate fusion of base and detail information
- Final fusion or reconstruction

Some additional measures can be taken within these three steps to further maximize the fusion rate. The first step of MSD is very important as it is used to decompose the input (source) images into finer and relatively coarser levels (Duan and Wang, 2021). The quality of decomposition and the effectiveness of the fusion mechanism used are both influenced by the MSD technique. For MSD either a transform or a spatial filter is used. By applying these transforms multiple scales of information are obtained as these transforms use orthonormal high-pass and low-pass filter banks to separate the information (Luo and Luo, 2023). The use of orthonormal filter banks helps in the exact reconstruction of the decomposed signal which is a vital step for transform-based decomposition (Veranyurt and Sakar, 2023). In the case of spatial domain-based decomposition, the image is first low-pass filtered using an existing filter kernel or a customized low-pass FIR or IIR filter and then high-pass information is extracted by subtracting the low-pass filtered image from the original image.

There are many such methods to decompose the image both in the transform domain and spatial domain. Some such methods in the transform domain are Discrete Cosine Transform, Wavelets and their shift-invariant variants such as Shift Invariant Wavelet Transform (SIWT), Dual-Tree Complex Wavelet Transform (DTCWT), Curvelets, Contourlets, Shearlets, etc. These transforms were developed in such order to obtain characteristics such as shift-invariance, multi-directional analysis capability faster computation speed, etc. Examples in the spatial domain are filters such as Average filter, Joint/Cross Bilateral filter, Gaussian filter, Bilateral filter, Guided filter, etc. These filters evolved in the same order to acquire better smoothing operation while preserving the edges and contours or preserving high-pass information (Bustos *et al.,* 2023).

In addition, the detail and base layers can be improved through some saliency detection tools or some other visibility enhancement technique. There exist broad ranges of these techniques which are investigated by researchers. Weights are generated using the salient object information which is extracted from the source images (Bhavana *et al.,* 2022). These weights are obtained either by normalizing the saliency maps or by filtering the saliency maps repeatedly using an appropriate edge-preserving filter. Those weights are then used to enhance the details. Finally, the base and detail layers can be merged using some appropriate fusion rule (Goyal *et al.,* 2021). There are some most popularly used fusion rules such as Choose max, choose min, Average fusion rule, Weighted average fusion rule, Fusion rule based on activity measurement, etc. The fusion rules are also called coefficient-selecting methods as they are used to select coefficients for the fused image out of multiple source images (Mahmoud, 2020).

Apart from evaluation with the eyes, there are various mathematical tools available to evaluate the quality of the fusion process. These tools can be categorized into two parts:

- No reference-based metrics
- Reference-based metrics

The no-reference-based methods do not require any image as a reference point to evaluate various properties of the fused image. On the other hand, reference-based metrics require reference images (Input source images) to evaluate the quality of the fused image. Earlier, no reference-based metrics like entropy, standard deviation, correlation, mutual information, and spatial frequency were quite popular, but nowadays, reference-based metrics like Q_{abf} (Fusion rate) are defined as in Eq. 1:

$$Q^{STF} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(Q^{SF}(i,j)\times\omega_S(i,j)\times\omega_T(i,j)))}{\sum_{i=1}^{M}\sum_{j=1}^{N}(\omega_S(i,j)+\omega_T(i,j)} \quad (1)$$

Where $Q^{ST}(i,j),=Q_{\beta}^{ST}(i,j)\times Q_{a}^{ST}(i,j),Q_{a}^{AB}(i,j)$ are edge strength and orientation preservation values, and $\omega_A(i,j),\omega_B(i,j)$ weights depict edge strength and preserved orientation values. N_{abf} (Number of artifacts) is defined as Eq. 2:

$$N^{STF} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} SM_{i,j}!(1-Q_{i,j}^{SF})w_{i,j}^{S}+\left(1-Q_{i,j}^{TF}\right)w_{i,j}^{T}\right]}{\sum_{i=1}^{M}\sum_{j=1}^{M}(w_{i,j}^{s}+w_{i,j}^{T})} \quad (2)$$

Where, $SM_{i,j}=\begin{cases}1\\0\end{cases}, if\ g_{i,j}^{F}\ ? > g_{i,j}^{S}\ and\ g_{i,j\,g_{i,j}^{F}}^{F}, g_{i,j}^{T}$ are the edge strengths of the fused image and the source images ST respectively $Q_{I,J}^{SF}$? And $Q_{I,J}^{TF}$ are gradient information

preservation parameters. $w_{i,j}^{S}$ and $w_{i,j}^{T}$ are corresponding weights of source images *S* and *T* respectively and L_{abf} (Loss of information) which can be calculated as in Eq. 3:

$$L_{abf} +=1-(Q_{abf}+N_{abf}) \quad (3)$$

An excellent review of these metrics along with some other assessment metrics is given by Liu *et al.* (2012).

Related Work

Image fusion managed to flourish in the last two decades due to its inherent advantages. The main advantages can be outlined as follows:

1. A compensation to the limits of various image-capturing sensors
2. Effective image content analysis due to added complementary information
3. Wide areas of applications like surveillance, remote sensing, concealed weapon detection, medical diagnosis through multi-modality image fusion, etc
4. Saving of storage space
5. Cost-effective as compared to hardware changes

Earlier, image fusion has been explored with some primitive state-of-the-art techniques like DCT, Laplace transform, Wavelets, etc. The main disadvantages encountered using those methods were the loss of information during Multi-scale Decomposition (MSD), shift-invariance, and limited directionality. However, these techniques are still used with some modifications to achieve better results.

Nowadays, a shift of focus happened toward hybrid techniques. Hybrid domain techniques are a communion of transform domain and spatial domain techniques that exploit the advantages of both domains. Non-subsampled MSD and salient object information extraction are important parts of these techniques. In light of this, we will compare our method with some recently proposed techniques based on Discrete Cosine Harmonic Wavelet Transform (DCHWT) by Liu *et al.* (2012), Singular Value Decomposition (SVD) by Naidu (2011), Guided filter (GF) by Shreyamsha (2013), Discrete Cosine Transform (DCT) by Li *et al.* (2013), Cross Bilateral filter (CBF) by Kaur *et al.* (2015), Saliency Detection 1 (SD1) by Shreyamsha (2015) Saliency Detection 2 (SD2) by Bavirisetti and Dhuli (2016), Anisotropic Diffusion (AD) by Bavirisetti *et al. (*2017), Fourth Order Partial Differential Equation (FPDE) by Zhan *et al.* (2017) and Fast filtering IF (FFIF) by Chen *et al.* (2019).

Naidu (2011) have proposed and evaluated an innovative fusion technique based on Multi-Resolution Singular Value Decomposition (MSVD). The performance of the algorithm is compared with the performance of a well-known wavelet-based picture fusion technique. MSVD image fusion has been demonstrated to be practically as effective as wavelets. The computational structure is fairly simple and may be suitable for real-time applications. In addition, in contrast to FFT, DCT, and wavelet, the basis vectors of MSVD are dataset-dependent and not predetermined.

Wavelets' multiresolution and energy compaction have made it possible for image fusion to merge vital information from source images, such as edges and textures, into one without introducing any artifacts to improve context and situational awareness (Shreyamsha, 2013). The wavelet transforms, being computationally expensive, are a representation of a convolution of the image in question by wavelet filter coefficients. The lifting-based wavelets have made calculations easier but at the cost of the performance and visual quality of the fused image. An image fusion based on Discrete Cosine Harmonic Wavelet (DCHWT) is proposed to maintain the performance and visual quality of the fused image at the cost of fewer calculations. The computational complexity of the proposed method is comparable to lifting-based wavelets and better than convolution-based wavelets.

An efficient and fast image fusion method is introduced by Li *et al.* (2013), for fusing a large number of photos to generate a highly informative fused image. This method is based on the 2-scale decomposition of an image into a detail layer that records small-scale details and a base layer that has large-scale intensity variations. For the ultimate utilization of spatial consistency for the merging of the detail and base layers, a novel weighted average approach grounded on guided filtering is proposed (Kaur *et al.,* 2015). Quality assessment measures for image fusion have been utilized to construct and research block DCT-based image fusion algorithms. Five designs for image fusion have been described, such as morphological DCT (MpDCT) derived from block DCT, subband DCT (SDCT), wavelet structure DCT (WSDCT), resizing DCT (RDCT), and feature DCT (FDCT). The image fusion algorithm from WSDCT has been revealed to be more efficient. Employing the calculated weights derived from the detailed images obtained by extracting them from the source images based on CBF, Shreyamsha (2015) proposed a weighted average fusing method for source photos. The performance of the method has been validated on some multisensor and multi-focus image pairs, and it has been visually and statistically compared to existing methods. It is found that for each performance measure, none of the methods has shown consistent performance. In most cases, the proposed method has worked optimally compared to them.

A new edge-preserving image fusion method for visible and infrared sensor images was introduced by

Bavirisetti and Dhuli *(*2016). The original images are decomposed into approximation and detail layers by anisotropic diffusion. Weighted linear superposition and the Karhunen-Loeve transform are applied to compute the final detail and approximation layers, respectively. The final detail and approximation layers are linearly fused to produce a fused image. Petrovic measures are employed to test the performance of the proposed algorithm.

A novel picture fusion method using fourth-order partial differential equations and principal component analysis is presented by Bavirisetti *et al.* (2017). Every source image is initially processed using fourth-order partial differential equations to generate detail and approximation images. Second, to find the optimal weights, detailed images are processed using principal component analysis. Third, these detailed images are combined using the help of suitable weights to create the final detailed image. Fourth, an average operation on approximation images is applied to create the final approximation image. Finally, the final approximation and detail images are merged to compute the resulting fused image. A picture fusion framework is proposed by Zhan *et al.* (2017) for different types of multimodal images with rapid spatial domain filtering. Contrast and sharpness are first sensed using the picture gradient magnitude. Then, the picture gradient magnitude is subjected to a rapid morphological closure process to close gaps. Third, a fast structure-preserving filter processes the weight map, which has been obtained from the multimodal image gradient magnitude. Finally, a weighted-sum rule is employed in constructing the fused image. The proposed method is faster than the fastest baseline algorithm by at most four times.

A fusion approach for cyber-physical systems is introduced by Chen *et al.* (2019). The MSMD-based technique is proposed as compared to the traditional approaches. Multi-scale decomposition is employed first to obtain detailed layers and basic layers with reserved edges. Secondly, rather than utilizing detailed layers as in the previous technique, multi-direction decomposition is employed to construct base layers. Subsequently, by choosing the maximum value based on the patch, a series of multidirectional base layers and detailed layers are created. A discrete wavelet transforms and pixel alignment image fusion technique is proposed by Mahmoud (2020). It is mainly applied to the detection of hidden weapons. Picture Wavelet (WT) and inverse wavelet transform (IWT) are employed in a data fusion approach that uses fusion dependence criteria for low-complexity sensors in terms of correlation coefficients. The coefficient with the highest correlation rate is selected as the fusion rule. The co-existing property is more prevalent as the correlation increases. The proposed method keeps higher quality while its real-time reaction speed is 40% faster compared to similar existing algorithms. It is superior to other algorithms in performance, with a better PSNR of more than 10% on average over similar algorithms. A new technique for detecting concealed firearms using DWT with the metaheuristic Harmony Search Algorithm and SVM classifier is proposed by Altaher *et al.* (2020). It creates a blended image by applying the hybrid Hotline transform, coupled with the standard discrete wavelets transform first. The optimal harmony is then determined by employing a heuristic search method. These are then grouped using K-means SVM to enhance classifiers for the discovery of hidden weapons.

Another image combination method is based on saliency finding and two-scale stands for picture decomposition proposed by Naidu *et al.* (2020). The approach is lucrative since saliency-based methods have been widely applied to the combination of visible (VIS) and infrared images, where visible and infrared images can store point-by-point basis information and emphasize the salient points' location at the same time. Another approach to generating weight maps based on visual saliency is proposed. The GAN framework was introduced by Yang *et al.* (2021) to reconstruct high-quality images from multi-source PMMWIs. The registration network is proposed for privacy issues and false alarm elimination, and the segmentation network employs multi-scale features to integrate global and local information in PMMWIs and visible images to acquire precise shape and position information in the images. The detection results of every individual frame are combined by a synthetic method. Based on experimental results, the proposed method has a fast detection rate, with 92.35 accuracy and 90.3% recall.

The Latent Low-Rank Method, an accurate image fusion technique for detecting hidden weapons or other objects under a person's clothing, was proposed by Bhavana *et al.* (2022). Salient features can be detected through latent low-rank representation. The rate of object detection is 94.6%. Fusion performance is evaluated subjectively based on numerous parameters. A system based on deep learning that is able to detect and find concealed guns on thermal images for real-time monitoring is introduced by Veranyurt and Sakar (2023). The concealed gun within the given thermal image is detected and found by a deep learning-based architecture. With a 0.84 F1 measure on the test set, an optimized VGG19-based model yielded the top test set outcomes in detecting the concealed handgun. The second module of the system attained the highest mean average precision value of 0F.95 in detecting and recognizing the gun within approximately 10 milliseconds due to an enhanced Yolo-V3 model.

To enhance the detection of faint thermal signals and the automation of infrared image categorization by Khor *et al. (*2024b) by combined IR imaging and CNNs with transfer learning. Real-time heat emissions from the

surface of human volunteers' garments were captured with a mid-wavelength infrared detector. Two image types, fuzzy-c-clustered and raw thermal images, were investigated in this regard. A holdout set was employed to provide receiver operating characteristic curves, showing that the models of raw and fuzzy-c clustered pictures had corresponding area-under-the-curves of 0.8934 and 0.9681, respectively.

Muñoz *et al.* (2025) introduced a two-step concealed handgun detection method. The approach initially identifies potential guns at the frame level and, subsequently, ensures that they belong to an identified person. To ensure accurate and reliable detection, alarms are triggered only under specific conditions. Precautionary alerts are established in case a weapon is detected, but no person is present. Wearable and mobile apps can be facilitated by an efficient method tailored for low-end embedded systems. Hands-free functionality is enabled by the deployment of the system on a chest-worn Android smartphone with a small thermal camera. Experimental results on our dataset show a mAP@50-95 of 64.52%, validating the effectiveness of the method.

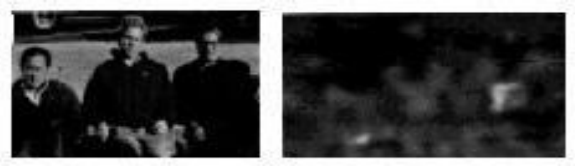
Materials and Methods

This section covers the details of the dataset, tools, and description of the proposed methodology.

Dataset Description

All the experiments have been performed on the "gun dataset," which is a challenging dataset captured in low-light conditions. This dataset is used to demonstrate the use of image fusion for concealed weapon detection. This can be extended to various other applications like surveillance and security.

The source images used are gray-scale images of size 200*256 and format ".gif". The first glance at the pictures shows that they are noisy and have artifacts at some boundaries, as shown in Fig. 1. The luminance and contrast of these images are also very low (Yang *et al.,* 2021). The image fusion will probably help in increasing the detail as well as the luminance of the overall image.



(a) (b)

Fig. 1: Source images (a) Visible image (b) IR image

Tools

The proposed method is based on Non-Subsampled Shearlet Transform (NSST) and Spectral residual saliency. The Shearlet transform is an extension of Wavelets. The basic transformation operations that can be applied to the matrices are scaling, rotation, translation, reflection, and shearing. These transformation operations, combined with filter kernels or certain analyzing functions, provide flexibility in the efficient analysis of images. In the Shearlet transform, an additional shearing function is applied along with the wavelet function, which already uses scaling and translation to get operations in multiple directions and hence achieve directionality. A normal horizontal shearing operation on a vector can be defined as shown in Eq. 4:

$$\begin{matrix} i' \\ j' \end{matrix} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \tag{4}$$

Where the parameter s decides the amount of backward or forward shear. Similarly, the continuous Shearlet transform has a dilation function formed with the product of shear matrices with the parabolic scaling matrices. The Shearlet coefficients can be obtained as shown in Eq. 5:

$$\varphi(b,s,t) = < x, \phi_{(bst)} > \tag{5}$$

Where, b>0 (scaling function),s, $s \in R$ (shear parameter), $t \in R^2$ (translation parameter) and x is the signal whose Shearlet coefficients are obtained as ϕ. To obtain a discrete representation, the three parameters are sampled at appropriate levels for optimum reconstruction. NSST is used in this technique, for the decomposition of the image because it provides the flexibility of analysis of the image in multiple directions. The conventional transforms like wavelet transform did not provide this flexibility. The wavelet transform could help with analysis in only 3 directions viz. Vertical, Horizontal, and Diagonal. This in turn helped in the efficient extraction of details from multiple directions hence improving the fusion rate.

Similarly, on the other hand, the spectral residual saliency provides salient information about objects by extracting spectral residue from the log spectrum of an image. In this method, first, the log spectrum of the image is obtained then the spectral residual is obtained by removing the statistically redundant information of the image. Then to create the object map, a thresholding operation is applied to the pixels by (Hou and Zhang 2007).

Methods

The general layout of the proposed method is:

Step 1: Multiscale decomposition using Shearlet transform.

1. Source image A, (I A) decomposed to I_B^{1A} (Base layer) and $\left\{I_D^{1A}, I_D^{2A}, I_D^{3A}, \cdots\cdots\cdots I_D^{NA}\right\}$ (Details) using NSST.
2. Source image B, (I_B) decomposed to I_B^{1B} (Base layer) and $\left[I_D^{1B}, I_D^{2B}, \cdots\cdots\cdots I_D^{2B}\right]$ (Details) using NSST.

Step 2: If one level of decomposition is chosen then go to step 3 otherwise add all the detail images to form a unified detail frame as in Eq. 6:

$$I_D^X = \sum_{i=1}^{n} I_D^{N_n} \tag{6}$$

Step 3: Simultaneously, extract the salient information from the source images as $S_A = \vartheta(I_A) S_B = \vartheta(I_B)$ and.

Note: Segmentation is an important part of the salient information extraction.

Step 4: Normalize the saliency maps to generate weights for the strengthening of the boundary pixels as $w_B = w_A = S_A / (S_A + S_B) w_B = S_B / (S_A + S_B)$ and.

Step 5: Calculate the weighted detail layers with the help of generated weights and then fuse them using the Choose max fusion rule by using Eq. 7:

$$I_D = maimum(w_A I_D^A, w_B I_D^B) \tag{7}$$

Step 6: The fused base layer is obtained by adding the corresponding visible and IR base layers as in Equation 8:

$$I_B = I_B^A + I_B^B \tag{8}$$

Step 7: Calculate the Inverse Shearlet transform on the final base and detail layers to obtain the final fused image.

After the final step fusion quality is evaluated using three metrics namely Fusion rate (Q_{abf}), Loss of information (L_{acf}), and artifact measure $(Q_{abf} N_{abf}$. These parameters follow the equation. So, the values of all three parameters lie within 0 and 1 and it is obvious from the terminology that the value of (Q_{abf}) should be close to 1 (ideally 1) and the values of $(Q_{abf} N_{abf}$ and should be close to zero (ideally 0).

Figure 2 shows the block diagram of the proposed methodology that utilizes a multi-modal image fusion approach based on the Shearlet transform for concealed weapon detection in low-light conditions. Source visual and infrared images are decomposed into base and detail layers using NSST. Salient features are extracted and normalized to generate weight maps for boundary enhancement, followed by max-rule fusion of detail layers and additive fusion of base layers. The final fused image is reconstructed using the inverse Shearlet transform. Fusion quality is assessed using Fusion Rate, Information Loss, and Artifact Measure, with optimal values targeting a high fusion rate and minimal loss/artifacts.

A visible image and an infrared (IR) image form the initial input images employed in the process. While the infrared image holds thermal data which can be utilized for finding out concealed objects, it often is not clear in terms of space. Conversely, the visible image possesses great spatial resolution and textural information. The proposed method initially employs a Non-Subsampled Shearlet Transform (NSST) to decompose every source image into a base layer, holding low-frequency information, and multiple detail layers, holding high-frequency features such as contours and edges, in an attempt to integrate the strengths of both modalities. Due to its shift-invariance and ability to store directional information in an array of orientations, both of which are necessary for structural integrity during fusion, the NSST is particularly beneficial in this case.

Simultaneously, the spectral residual method, which partitions the most prominent features of an image by discarding redundant spectral information from its log spectrum, is employed for developing the saliency maps of both input images. The pixel-wise weight maps obtained from normalizing these saliency maps assign more weight to the regions with more prominent visual or thermal information, thereby guiding the fusion process. By employing a weighted Choose-Max process, the detail layers from both modalities are combined, retaining at every pixel location the highest saliency weight-normalized detail coefficient. The most informative elements from both input sources are retained as a consequence.

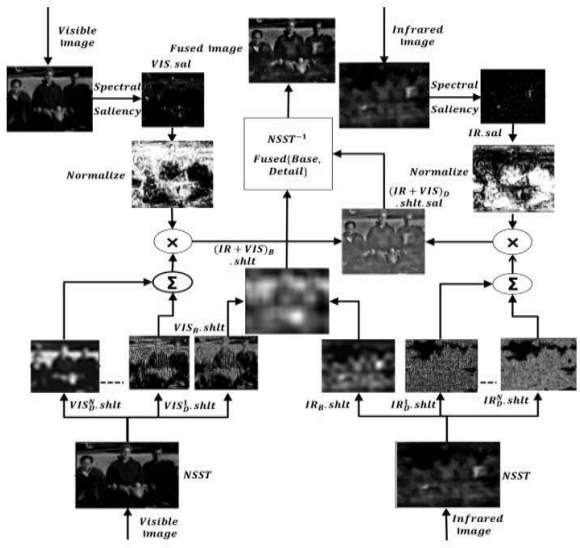


Fig. 2: Block diagram of the proposed methodology

A simpler additive fusion algorithm is applied for the base layers, where the base layers from the visible and infrared images are summed to form a single base. The full fused image is reconstructed by recombining the fused base and detail layers through the inverse NSST. It is expected that this output will demonstrate better representation of hidden objects or weapons, better clarity, and better contrast. There are three measures employed to measure the performance of the fusion: Fusion Rate (Qavz), which estimates the extent to which the two source images are incorporated; Loss of Information (L), measuring the level of content lost in the process of fusion; and Artifact Measure (Na), indicating the level of unwanted distortions produced while performing the process.

Results and Discussion

This section covers the results that are carried out by the experiments of the proposed methodology and also shows the comparison with the existing techniques.

Figure 3 shows the visual results for various techniques. The boxes highlight the parts to be observed. It is quite evident that techniques like DCHWT, GF, CBF, SD 1 and 2 introduce some artifacts as seen in the upper highlighting block. The results obtained from techniques like SVD, DCT, AD, and FPDE are hazy, and in fast-filtering IF and SD 2, the concealed weapon is almost invisible. The results that are carried out from the proposed method are very close to the characteristics of the input source images, because of this reason this technique provides better objective evaluation results.

Table 1 demonstrates the comparative analysis of the proposed method with existing techniques. The proposed method surpasses all current methods with the highest fusion rate of 0.9276. Although they perform well, methods such as CBF (0.9024), FFIF (0.8991), and DCHWT (0.8890) lag behind the proposed method by a slight margin. Compared to other methods, SVD (0.6129) has the lowest fusion rate, indicating that it retains much less information. with the minimum information loss (0.0536). DCHWT (0.1107) and CBF (0.0953) also show good performance. The worst approach in this situation is SVD (0.3864) since it suffers from maximum information loss. The proposed method shows an artifact (0.0128), indicating nearly no extra noise or visual distortion. SVD (0.0007) and DCHWT (0.0003) both show high artifact reduction ability. GF (0.0100) and CBF (0.0363) have relatively higher artifact values, meaning that they introduce more distortions.

The efficiency of integrating information from visible and infrared source images into the fused image is quantified by the Fusion Rate. Higher fusion quality and better knowledge retention are reflected by a higher fusion rate (closer to 1). As shown in Fig. 4, the proposed method surpasses all current methods with the highest fusion rate of 0.9276. Although they perform well, methods such as CBF (0.9024) and DCHWT (0.8890) lag behind the proposed method by a slight margin. Compared to other methods, SVD (0.6129) has the lowest fusion rate, indicating that it retains much less information. The proposed technique generates the most complete fused image by combining useful information from both sources in an optimal way.

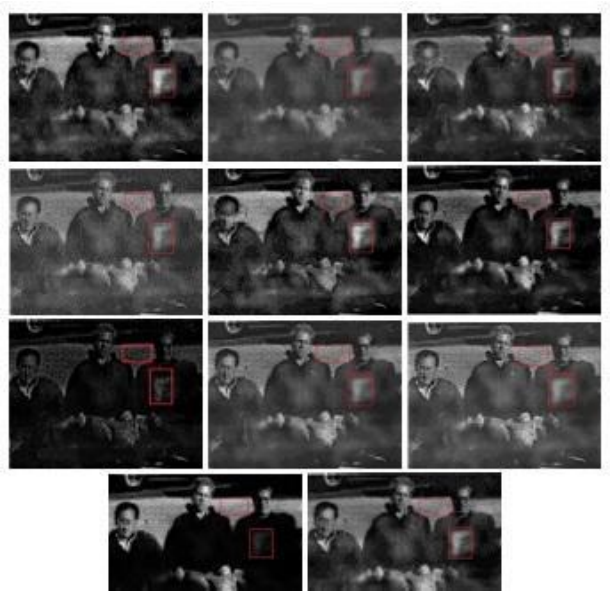


Fig. 3: Image fusion results using First row (Left to right): DCHWT, SVD, GF, second row (Left to right): DCT, CBF, Saliency detection 1, third row (Left to right): Saliency detection 2, Anisotropic diffusion, FPDE, fourth row (Left to right): Fast filtering IF, Proposed method

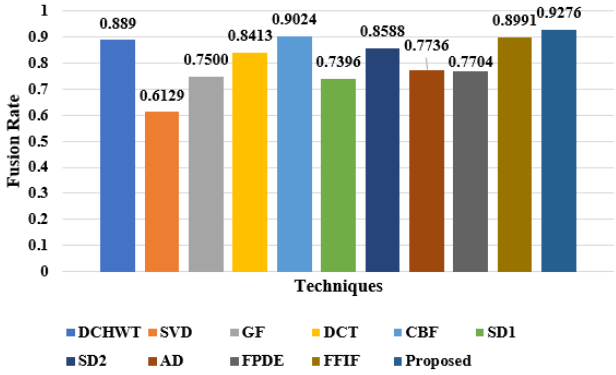


Fig. 4: Fusion rate comparison of the Proposed model with existing techniques

Table 1: Comparative Analysis of Proposed Method with Existing Techniques

Techniques	Fusion Rate	Loss of Information	Artifact Amount
DCHWT	0.8890	0.1107	0.0003
SVD	0.6129	0.3864	0.0007
GF	0.75	0.2400	0.0100
DCT	08413	0.1561	0.0026
CBF	0.9024	0.0953	0.0363
SD1	0.7396	0.2600	0.0004
SD2	0.8588	0.1367	0.0045
AD	0.7736	0.2236	0.0001
FPDE	0.7704	0.2291	0.0005
FFIF	0.8991	0.0993	0.0016
Proposed	0.9276	0.0596	0.0128

The Loss of Information metric measures the quantity of original data lost in the fusion process. Most of the original information is preserved when the values are low (closer to 0). Almost all-important information is maintained by the proposed method as shown in Fig. 5, which shows the minimum information loss (0.0536). DCHWT (0.1107) and CBF (0.0953) also show good performance. The worst approach in this situation is SVD (0.3864) since it suffers from maximum information loss. Again, the proposed method avoids information loss throughout fusion, making it superior.

The quantity of unwanted distortions or noise added during the fusion process is quantified by the artifact amount. For clean and realistic photos, lower artifact values (closer to 0) are optimal. As shown in Fig. 6, the proposed method shows artifact (0.0128), indicating nearly no extra noise or visual distortion. SVD (0.0007) and DCHWT (0.0003) both show high artifact reduction ability. GF (0.0100) and CBF (0.0363) have relatively higher artifact values, meaning that they introduce more distortions.

Computational Considerations and Limitations

The Non-Subsampled Shearlet Transform (NSST) and the extraction of spectral residual saliency are the two dominant parameters that influence the computing efficiency of the proposed fusion framework. Even though NSST provides shift-invariant and multi-directional analysis, its application of directional filtering on many scales and the absence of subsampling increase its computational cost over basic wavelet or pyramid-based transform. However, the parallelizable operations of the algorithm do make it suitable for CPU versions with optimization or GPU acceleration. Running two 200x256 visible and infrared images using MATLAB to combine them takes approximately 2.1 seconds on average per set of images on a standard i5 processor with 8GB RAM. For offline surveillance analysis, this is acceptable, but for real-time use, it may have to be optimized. The spectral residual method is light on computations in terms of saliency extraction, involving only an inverse Fourier transform and a log-spectrum calculation. It is a good choice for real-time attention modelling due to its simplicity, which ensures that it does not significantly add to the runtime.

In addition, in processing high-resolution images or video streams, memory overhead associated with NSST can increase dramatically as the number of directional components or levels of decomposition increases. Due to this, finding a balance for the depth of decomposition with real-time constraints is essential. Scalability is also a concern; while the present system works decently on smaller image files, it would require additional memory management and parallelization strategies for larger data sets or batch processing of multiple camera feeds.

Latency constraints and technological boundaries also need to be considered in practical applications such as intelligent traffic monitoring or airport surveillance. As an instance, application of this method to edge devices or embedded platforms would require utilization of model compression algorithms or light-weight approximations. In addition, fidelity of saliency map production and fusion can be affected by variations in ambient lighting, occlusion, and atmospheric noise between scenes. Thus, one of the most important directions for future growth is robustness against a range of environmental conditions.

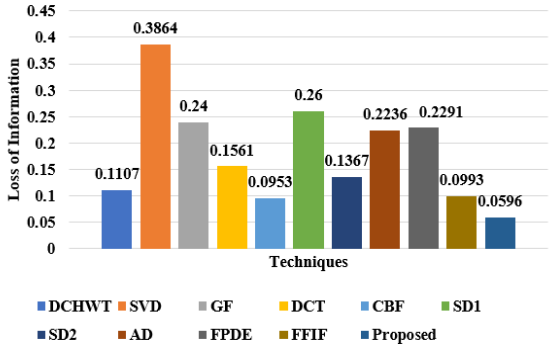


Fig. 5: Loss of information comparison of the Proposed model with existing techniques

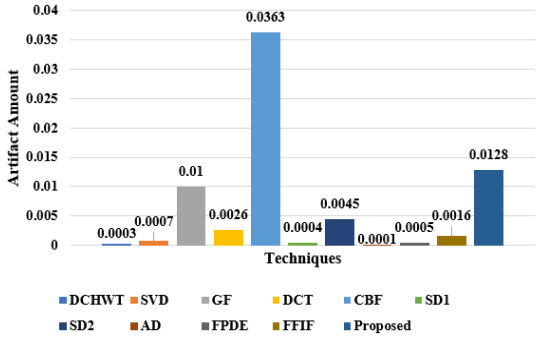


Fig. 6: Artifact comparison of the proposed model with existing techniques

Despite these advantages, the proposed method has some disadvantages. To date, it is dependent on input visible and infrared images being pre-registered; misalignment between modalities may decrease the precision of the fusion. Additionally, shear direction selection and decomposition levels in NSST affect runtime and quality of results, and optimal tuning for different datasets would be required. Also, the solution is not adaptive; rather than applying learning-based decision-making mechanisms, it utilizes fixed principles (choose-max and additive fusion), which could limit performance in noisy or extremely dynamic environments. To enhance fusion robustness and generalizability to a range of surveillance environments, future research will focus on real-time deployment using GPU-based acceleration and the integration of learning-based adaptive weighting mechanisms.

Subsequent studies will attempt to integrate dynamic parameter selection processes, where the number of scales, shearing parameters, and fusion weights are dynamically tuned based on scene complexity or saliency detection confidence levels, to mitigate these problems. There can be further reduction in dependence on manually designed thresholds and fusion logic by using machine learning models for saliency prediction and fusion rule optimization. Ultimately, these improvements will contribute to the development of a fusion system that is more scalable, real-time capable, and context-aware, and fit for large-scale deployment in security-critical applications.

Implications

The suggested fusion technique, which makes use of spectral residual saliency and Non-Subsampled Sharlet Transform, has direct implications for improving the accuracy and dependability of surveillance systems, particularly in dimly lit or obscured areas like airports, high-security zones, and public transportation hubs. The technique may greatly enhance concealed weapon detection capabilities by skill fully integrating thermal and visual spectrum data, which will enhance public safety and threat prevention. Additionally, the framework can be expanded to other crucial areas, such as military reconnaissance, wildlife monitoring, and autonomous driving in nighttime or foggy settings. For even more automation and accuracy, its modular design enables integration with deep learning models. The creation of low-cost, real-time, embedded image fusion systems that may be deployed on edge devices is made possible by this study.

Moreover, the versatility of this fusion strategy allows it to be integrated into Internet of Things (IoT)-based security systems and multi-sensor platforms, wherein real-time visual understanding is critical. Data fusion among several imaging modalities may enable proactive surveillance, intelligent alarm systems, and enhanced situational awareness as smart cities evolve. This method may be adapted for application in medicine to integrate multi-modal medical images, e.g., visible light and infrared thermography, to enhance screening and diagnosis in resource-constrained environments.

Furthermore, the proposed methodology provides the core framework for academic research and development in computer vision and computational imaging. It provides an imitable and scalable template for future advancement by combining modern saliency modelling and performance-based evaluation metrics with conventional signal processing methods (like NSST). This study may be employed as a pedagogical case study in schools by instructors and students who are interested in transform-domain image fusion and its potential uses.

The application of intelligent, self-optimizing fusion models on embedded devices such as drones, intelligent security cameras, and mobile phones becomes not just possible but also vital as AI and edge computing capacities continue to evolve. Thus, the broader implications of this study extend far beyond the detection of hidden weapons; it also addresses the growing requirement for intelligent, context-sensitive vision systems across multiple industries, such as environmental monitoring and security.

Conclusion

In this article, a new paradigm based on NSST for the fusion of a challenging multi-sensor dataset for concealed weapon detection is proposed. Due to the well-known fact that the NSST coefficients exhibit the dependencies relationship of inter-direction, interscale, and also between the neighbors, an efficient image fusion algorithm is proposed. This is necessary as an image has region boundaries or contours, not restricted to any particular direction. The use of spectral saliency detection further facilitated in transfer of necessary target information from the source images. This technique is tested on various multi-sensor image pairs. The results for this particular dataset are compared with various primitive and recently proposed, high-performance techniques, and it has been found that it presents better results in terms of objective and subjective evaluation. Experimental analysis explicitly indicates that the proposed technique can address the issue of mismatch between subjective and objective analysis. However, the problem of computational complexity is also resolved by the proposed method. The proposed method surpasses all current methods with the highest fusion rate of 0.9276, minimum information loss (0.0536), and shows artifact (0.0128), indicating nearly no extra noise or visual distortion. Among all the methods that are discussed, the proposed method is most effective as it achieves the ideal

balance among controlled artifacts, minimal information loss, and optimal information fusion. In our Upcoming work, we will devise more efficient integration strategies to further enhance the efficacy of the algorithm performance. The primary motive is to promote image integration techniques in remote sensing and surveillance applications, along with addressing the critical issues of data pre-processing and image alignment probes as well.

Acknowledgment

We thank Anupma Gupta (Department of Electronics and Communication Engineering, Chandigarh University, Punjab, India) and Vinay Kukreja (Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India) for supervising and guiding us throughout the revision process, for carefully reviewing and revising the manuscript in response to the journal's reviews.

Funding Information

The authors received no financial support for this article's research, authorship, and/or publication.

Authors Contributions

Apoorav Sharma: Concept, experimentation, writing and edited.

Shagun Sharma and Kalpna Guleria: Methodology, formal analysis and validation.

Ayush Dogra: Experimentation and writing.

Pankaj Lathar: Supervision, analysis and validation of results, edited and proofread.

Archana Saini: Analysis of results, review and proofread.

Bhawna Goyal: Investigation, validation and visualization of results.

Conflict of Interest

The authors do not have any conflicts of interest.

References:

Altaher, A. W., & Hussein, A. H. (2020). Intelligent security system detects the hidden objects in the smart grid. *Indonesian Journal of Electrical Engineering and Computer Science*, *19*(1), 188–195. https://doi.org/10.11591/ijeecs.v19.i1.pp188-195

Bavirisetti, D. P., & Dhuli, R. (2016). Fusion of Infrared and Visible Sensor Images Based on Anisotropic Diffusion and Karhunen-LoeveS Transform. *IEEE Sensors Journal*, *16*(1), 203–209. https://doi.org/10.1109/jsen.2015.2478655

Bavirisetti, D. P., Xiao, G., & Liu, G. (2017). Multi-sensor image fusion based on fourth order partial differential equations. *Proceedings of the 2017 20th International Conference on Information Fusion (Fusion)*, 1–9. https://doi.org/10.23919/icif.2017.8009719

Bhavana, D., Kishore Kumar, K., & Ravi Tej, D. (2022). Infrared and visible image fusion using latent low rank technique for surveillance applications. *International Journal of Speech Technology*, *25*(3), 551–560. https://doi.org/10.1007/s10772-021-09822-2

Bustos, N., Mashhadi, M., Lai-Yuen, S. K., Sarkar, S., & Das, T. K. (2023). A systematic literature review on object detection using near infrared and thermal images. *Neurocomputing*, *560*, 126804. https://doi.org/10.1016/j.neucom.2023.126804

Chen, L., Yang, X., Lu, L., Liu, K., Jeon, G., & Wu, W. (2019). An image fusion algorithm of infrared and visible imaging sensors for cyber-physical systems. *Journal of Intelligent & Fuzzy Systems*, *36*(5), 4277–4291. https://doi.org/10.3233/jifs-169985

Dey, S., & Aravind, A. O. (2024). *Review of Concealed Object Detections Using RF Imaging Over UWB Regime for Security Applications*.

Dong, L., & Wang, J. (2024). Fusion CPP: Cooperative fusion of infrared and visible light images based on PCNN and PID control systems. *Optics and Lasers in Engineering*, *172*, 107821. https://doi.org/10.1016/j.optlaseng.2023.107821

Duan, C., & Wang, Z. (2021). Infrared and Visible Image Fusion Using Multi-Scale Edge-Preserving Decomposition and Multiple Saliency Features. *Optik – International Journal for Light and Electron Optics*, *22*(8), 165775. https://doi.org/10.1016/j.ijleo.2020.165775

Goyal, B., Dogra, A., Khoond, R., Gupta, A., & Anand, R. (2021). Infrared and Visible Image Fusion for Concealed Weapon Detection using Transform and Spatial Domain Filters. *Proceedings of the 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) – ICRITO*, 1–6. https://doi.org/10.1109/icrito51393.2021.9596074

Hadinejad, I., Amiri, M. A., & Fahimifar, M. H. (2025). Passive millimeter wave and visible image fusion using concealed object detection and gradient transform. *Signal, Image and Video Processing*, *19*(2), 181. https://doi.org/10.1007/s11760-024-03761-6

Hou, X., & Zhang, L. (2007). Saliency Detection: A Spectral Residual Approach. *Proceedings of the 3rd 2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. https://doi.org/10.1109/cvpr.2007.383267

Jagalingam, P., & Hegde, A. V. (2007). Pixel Level Image Fusion—A Review on. *Proceedings of the 3rd World Conference on Applied Sciences, Engineering and Technology (WCSET 2014)*, 8–12.

Kaur, H., KoundaI, D., & Kadyan, V. (2015). Image Fusion Techniques: A Survey. *Archives of Computational Methods in Engineering*, *28*(7), 4425–4447. https://doi.org/10.1007/s11831-021-09540-7

Khor, W., Chen, Y. K., Roberts, M., & Ciampa, F. (2024a). Non-contact, portable, and stand-off infrared thermal imager for security scanning applications. *AIP Advances*, *14*(4), 045314. https://doi.org/10.1063/5.0188862

Khor, W., Chen, Y. K., Roberts, M., & Ciampa, F. (2024b). *Infrared thermography as a non-invasive scanner for concealed weapon detection*. https://doi.org/10.17862/cranfield.rd.25028030.v2

Li, S., Kang, X., & Hu, J. (2013). Image Fusion With Guided Filtering. *IEEE Transactions on Image Processing*, *22*(7), 2864–2875. https://doi.org/10.1109/tip.2013.2244222

Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R., & Wu, W. (2012). Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(1), 94–109. https://doi.org/10.1109/tpami.2011.109

Luo, Y., & Luo, Z. (2023). Infrared and Visible Image Fusion: Methods, Datasets, Applications, and Prospects. *Applied Sciences*, *13*(19), 10891. https://doi.org/10.3390/app131910891

Mahmoud, H. A. H. (2020). A Novel Image Fusion Scheme using Wavelet Transform for Concealed Weapon Detection. *International Journal of Advanced Computer Science and Applications*, *11*(2). https://doi.org/10.14569/ijacsa.2020.0110239

Muñoz, J. D., Ruiz-Santaquiteria, J., Deniz, O., & Bueno, G. (2025). Concealed Weapon Detection Using Thermal Cameras. *Journal of Imaging*, *11*(3), 72. https://doi.org/10.3390/jimaging11030072

Naidu, A. R., Bhavana, D., Revanth, P., Gopi, G., Kishore, M. P., & Venkatesh, K. S. (2020). Fusion of visible and infrared images via saliency detection using two-scale image decomposition. *International Journal of Speech Technology*, *23*(4), 815–824. https://doi.org/10.1007/s10772-020-09755-2

Naidu, V. P. S. (2011). Image Fusion Technique using Multi-resolution Singular Value Decomposition. *Defence Science Journal*, *61*(5), 479. https://doi.org/10.14429/dsj.61.705

Shreyamsha, B. K. (2013). Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform. *Signal, Image and Video Processing*, *7*(6), 1125–1143. https://doi.org/10.1007/s11760-012-0361-x

Shreyamsha, B. K. (2015). Image fusion based on pixel significance using cross bilateral filter. *Signal, Image and Video Processing*, *9*(5), 1193–1204. https://doi.org/10.1007/s11760-013-0556-9

Veranyurt, O., & Sakar, C. O. (2023). Concealed pistol detection from thermal images with deep neural networks. *Multimedia Tools and Applications*, *82*(28), 44259–44275. https://doi.org/10.1007/s11042-023-15358-1

Wang, S., Du, Y., Lin, J., Zhao, S., & Dong, G. (2024). Infrared and visible military image fusion strategies and applications based on composite decomposition and multi-fuzzy theory. *Research Square*, *81*(12), 1–12. https://doi.org/10.21203/rs.3.rs-4721382/v1

Yang, H., Zhang, D., Qin, S., Cui, T. J., & Miao, J. (2021). Real-Time Detection of Concealed Threats with Passive Millimeter Wave and Visible Images via Deep Neural Networks. *Sensors*, *21*(24), 8456. https://doi.org/10.3390/s21248456

Zhan, K., Xie, Y., Wang, H., & Min, Y. (2017). Fast filtering image fusion. *Journal of Electronic Imaging*, *26*(06), 063004. https://doi.org/10.1117/1.jei.26.6.063004

Zhu, H., & Zhang, W. (2025). Infrared and Visible Image Fusion Based on Image Enhancement and Target Extraction. *IEEE Access*, *13*, 61862–61875. https://doi.org/10.1109/access.2025.3557799