Enhanced infrared and visible image fusion using a fast guided filter and an improved visual saliency map

Phu-Hung Dinh[1]

Abstract: The fusion of infrared (IR) and visible images aims to generate a composite image that is both highly informative and visually optimized for human perception or computer vision applications. This paper presents an innovative fusion approach designed to enhance the effectiveness of this process. The method begins by utilizing the Fast Guided Filter (FGF) to decompose input images into base and detail components. An improved visual saliency map technique is introduced to efficiently fuse the base components, while a local energy-based fusion strategy is applied to the detail components. Experimental evaluations demonstrate that the proposed approach outperforms conventional methods, delivering superior results in both subjective and objective assessments.
Keywords: Fast guided filter (FGF), visual saliency map (VSM), image fusion (IF).

1. Introduction

Image fusion (Ma, Ma, et al., 2019) is a specialized field within information fusion that leverages the diversity of information to enhance quality of image. Different imaging detectors capture images with unique characteristics, emphasizing distinct aspects of a scene. For instance, infrared detectors rely on the target's thermal radiation, producing images with lower resolution and less detail compared to visible images. Conversely, visible detectors capture images based on variations in spectral reflections, resulting in high-resolution images rich in texture and detail. The goal of image fusion is to combine the critical features from both types of images into a single, comprehensive representation, offering robust support for subsequent image analysis and interpretation. The imaging system of visible light detectors excels in delivering high-resolution images, capturing target edges, textures, and details with remarkable clarity. However, its ability to resist interference is limited, making it susceptible to disruptions caused by natural factors such as adverse or complex weather conditions. The infrared detector imaging system relies on the radiation characteristics of targets to generate images. Targets with higher temperatures emit stronger infrared radiation, resulting in higher grayscale values in the image, which enhances their prominence. As a form of passive imaging, infrared imaging does not require sunlight, offers excellent concealment, and can penetrate obstacles like smoke, bushes, and grass to capture images. It is also largely unaffected by ambient light, enabling effective imaging both day and night. These advantages make infrared detectors particularly well-suited for identifying people, animals, and military targets in concealed environments such as grasslands and forests. Consequently, infrared imaging technology is gaining significant attention in target monitoring systems. However, infrared images face challenges such as low resolution, poor contrast, and limited clarity, which can diminish their visual quality. To address these limitations, the fusion of infrared and visible light images through feature-based information integration offers an effective solution. This fusion not only enhances the complementarity of image data but also broadens the scope of detection, providing improved image quality and more comprehensive monitoring capabilities.

To date, approaches for solving the problem of Infrared (IR) and visible image fusion can be broadly classified into two categories: traditional methods and deep learning-based methods (Meher et al., 2019). Traditional methods typically follow a three-step process. First, the input images are decomposed into base and detail components. Second, these components are fused separately. Finally, an inverse transformation is applied to recombine the fused components into the final composite image. Image decomposition in these methods often leverages multi-scale transformation techniques or edge-preserving filters (S. Li et al., 2013) to achieve effective separation. To fuse the base and detail components, several commonly used methods include the max selection method, the averaging method, and methods based on local energy functions. Deep learning-based methods, on the other hand, can be divided into two subcategories (Yang et al., 2024). The first involves deep learning networks that play a partial role in the fusion process, focusing on either the base or detail components. The second employs end-to-end deep learning models that handle the entire fusion process holistically, streamlining the task from input to output.

According to our observations, the base component fusion method based on the VSM was proposed by (Ma et al., 2017). Since the VSM method focuses on regions with high contrast or areas containing significant information, it enhances the quality of the base component during fusion. Furthermore, when

[1]Thuyloi University
Received 10th Feb. 2025
Accepted 3rd Apr. 2025
Publication date 31st Dec. 2025

applied to base component fusion, VSM improves sharpness and detail in critical regions. However, the VSM method primarily emphasizes local regions, which may lead to the neglect or degradation of important global features, particularly in low-contrast base regions. Additionally, VSM can be affected by noise in the input data, potentially emphasizing unimportant regions or overlooking crucial information. To enhance the effectiveness of the VSM, we propose integrating global and local information to better highlight important regions while maintaining a balance between local details and overall structure.

The following summarizes some of our key contributions:

- First, we introduce an enhanced VSM method, named IVSM, to improve saliency detection.
- Second, we develop a base component fusion strategy, leveraging IVSM, termed FB_IVSM, to enhance structural preservation.
- Third, we propose a detail component fusion approach based on Distance-Weighted Regional Energy (DWRE) to refine texture and edge details.

The rest of the paper is organized into four sections. Section 2 presents the necessary background, covering VSM, FGF, and DWRE. Section 3 introduces the proposed IVSM method along with the image fusion approach for both base and detail components. Section 4 details the experimental setup and results. Finally, Section 5 concludes the paper and discusses potential directions for future research.

2. Background

2.1. Visual Saliency Map (VSM)

The VSM (Ma et al., 2017) is a method used in image processing to highlight regions of interest within an image. It emphasizes areas with high contrast or significant information. The saliency of a pixel p is defined based on the contrast of its intensity with all other pixels in the image:

$$V(p) = \sum_{i=1}^{N} \left| I(p) - I(i) \right| \quad (1)$$

V(p): saliency value of pixel p. I(p): intensity value of pixel p. I(i): intensity value of other pixels in the image I. N: total number of pixels in the image I.

To improve efficiency, the saliency value is computed based on the histogram of the image, rather than all N pixels.

$$V(p) = \sum_{j=1}^{L-1} M_j \left| I(p) - I_j \right| \quad (2)$$

L: number of intensity levels (256 for an 8-bit image). M_j: number of pixels with intensity level j. I_j: intensity value of level j.

2.2. Fast Guided Filter (FGF) and Image Decomposition

2.2.1. Fast Guided Filter

The Fast Guided Filter (FGF) (He & Microsoft, 2015) is an accelerated version of the Guided Filter, which is widely used for edge-preserving image filtering. It offers high-quality smoothing, preserves edges, and is computationally efficient. The guided filter is based on a local linear model, allowing it to suppress gradient-reversal artifacts and produce visually pleasing results. The fast variant significantly reduces computational complexity, making it suitable for high-resolution images and real-time applications.

Input: Input Image (p), Guidance Image (I), Filter Radius (r), Regularization Parameter (ϵ), Subsampling Ratio (s).

Output: Filtered Image (q).

Step 1: Subsample the Images.

Downsample the guidance image I and input image p by a factor of s according to Eqs. (3) and (4).

$I' = \text{subsample}(I, s)$ (3)

$p' = \text{subsample}(p, s)$ (4)

Adjust the filter radius proportionally: r' = r / s

Step 2: Compute Local Means and Correlations.

Adjust the filter radius proportionally according to Eqs. (5) and (6).

$m_I = \text{mean}(I', r')$ (5)

$m_p = \text{mean}(p', r')$ (6)

Compute the correlation terms according to Eqs. (7) and (8).

$C_I = \text{mean}(I' \circ I', r')$ (7)

$C_{Ip} = \text{mean}(I' \circ p', r')$ (8)

Step 3: Compute Variance and Covariance.

Variance of I′ in each local window according to Eq. (9).

$V_I = C_I - m_I \circ m_I$ (9)

Covariance between I′ and p′ in each local window according to Eq. (10).

$Cov_{Ip} = C_{Ip} - m_I \circ m_p$ (10)

The symbol ∘ denotes the Hadamard product operator.

Step 4: Compute Linear Coefficients

Calculate the linear coefficients a and b for each local window according to Eqs. (11) and (12).

$a = Cov_{Ip} / (V_I + \epsilon)$ (11)

$b = m_p - a \circ m_I$ (12)

Step 5: Upsample Coefficients

Upsample a and b to match the original resolution according to Eqs. (13) and (14).

$a_1 = upsample\ (a, s)$ (13)

$b_1 = upsample(b, s)$ (14)

Step 6: Compute Final Output

Use the upsampled coefficients and the original guidance image I to compute the output according to Eq. (15).

$q = a_1 \circ I + b_1$ (15)

2.2.2. Image Decomposition

An image decomposition method based on FGF (called ID_FGF) is proposed. The steps of the algorithm are described as follows:

Input: An image I and essential parameters.

Output: Base component B and detail component D.

Step 1: The FGF is employed on the input image I, resulting in a base component B.

Step 2: The detail component is computed according to Eq. (16).

$D = I - B$ (16)

2.3. Distance-Weighted Regional Energy (DWRE)

The DWRE technique (Panigrahy et al., 2023) enhances low-pass sub-bands in medical image fusion by weighting neighboring pixels based on their distance from the central pixel. Closer pixels receive higher weights due to their stronger influence on local structure, while distant ones contribute less. This ensures a more balanced and precise fusion of low-pass sub-bands. The DWRE of the detail component D in the input image I is determined using Eq. (17).

$$DWRE(i,j) = \sum_{k=-1}^{1} \sum_{l=-1}^{1} W(k+2, l+2)(D(i+k, j+l))^2 \quad (17)$$

Here, (i, j) denotes an image pixel, while W is a (3 x 3) matrix that assigns weights to neighboring coefficients, defined as per Eq. (18).

$$W = \begin{bmatrix} \frac{1}{1+\sqrt{2}} & \frac{1}{2} & \frac{1}{1+\sqrt{2}} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{1+\sqrt{2}} & \frac{1}{2} & \frac{1}{1+\sqrt{2}} \end{bmatrix} \quad (18)$$

3. The proposed model

3.1. Improved Visual Saliency Map (IVSM)

From Eq. (2) used to calculate the traditional V(p), we observe the following limitations: First, there is no spatial relationship between pixels. The formula only considers the intensity contrast between the pixel I(p) and the gray levels I_j, without taking into account the spatial distance between pixel p and the regions containing I_j. As a result, regions with the same intensity values are treated equally, regardless of their spatial positions in the image. Second, the formula has high sensitivity to noise. The formula is highly sensitive to noise in the image because the intensity values I(p) and I_j may not accurately represent actual salient features. If the image contains noisy pixels with abnormal intensities, the histogram-based weighting M_j can become unreliable.

Therefore, the proposed idea for the improved formula V(p) (referred to as IVSM(p)) introduces spatial weighting to overcome the limitations of the traditional formula. The main concept is to enhance the computation of V(p) by incorporating spatial context, which the original formula neglects. Here's the rationale:

Incorporation of Spatial Relationships:

- By adding a spatial weighting factor, the formula considers the Euclidean distance d(p, j) between the pixel p and the regions containing intensity I_j.
- This ensures that regions with the same intensity values I_j, but located in different spatial positions, are treated differently. This resolves the limitation where the original method ignored spatial relationships.

Control of Spatial Influence:

- The Gaussian function is introduced to model how the influence of a region diminishes with increasing spatial distance.
- The parameter σ adjusts how strongly the spatial distance impacts the weighting, allowing flexibility in controlling the spatial influence.

Add spatial weighting based on the Euclidean distance between pixel p and the regions containing intensity I_j to introduce spatial context according to Eq. (19).

$$V(p) = \sum_{j=1}^{L-1} M_j \left| I(p) - I_j \right| exp\left(- \frac{d(p,j)^2}{2\sigma^2} \right) \quad (19)$$

- d (p, j): The spatial distance between pixel p and the region with intensity j.

- σ: A parameter to control the spatial influence.

3.2. The base component fusion method using IVSM (FB_IVSM)

In this subsection, we introduce the fusion method for base components based on IVSM as follows:

Input: Two input images I_1 and I_2, two base components B_1 and B_2.
Output: Fused base component B_F.

Step 1: Applying the IVSM method to the two input images I_1 and I_2, we obtain two maps M_1 and M_2.
Step 2: Calculate the weight maps W_1 and W_2 (Ma et al., 2017) based on the two maps M_1 and M_2 using Eq. (20) and Eq. (21).

$W_1 = 0.5 + (M_1\text{-}M_2)/2$ if $M_1 > M_2$ (20)

$W_2 = 0.5 + (M_2\text{-}M_1)/2$ if $M_1 < M_2$ (21)

Step 3: Calculate the fused component B_F using Eq. (22), where ◦ denotes the Hadamard product operator.

$B_F = W_1 \circ B_1 + W_2 \circ B_2$ (22)

3.3. The detail component fusion method using DWRE (FD_DWRE)

In this subsection, we present the DWRE-based fusion method (Panigrahy et al., 2023) for detail components as follows:

Input: Two detail components D_1 and D_2.
Output: Fused detail component D_F.

Step 1: Applying the DWRE method to the detail components D1 and D2, we obtain the components E_1 and E_2.
Step 2: Calculate the fused map M using Eqs. (23) and (24).

$M_1(i,j) = 1$ nếu $|E_1| \geq |E_2|$ (23)

$M_2(i,j) = 0$ nếu $|E_2| < |E_1|$ (24)

Step 3: Apply Gaussian filtering to smooth the maps M_1 and M_2.
Step 4: Compute the fused detail component D_F using Eq. (25), where ◦ denotes the Hadamard product operator.

$D_F = M_1 \circ D_1 + M_2 \circ D_2$ (25)

3.4. Proposed image fusion method

The proposed image fusion method, named IVSM_DWRE, is illustrated in Figure 1. The steps of the proposed fusion algorithm are described as follows:

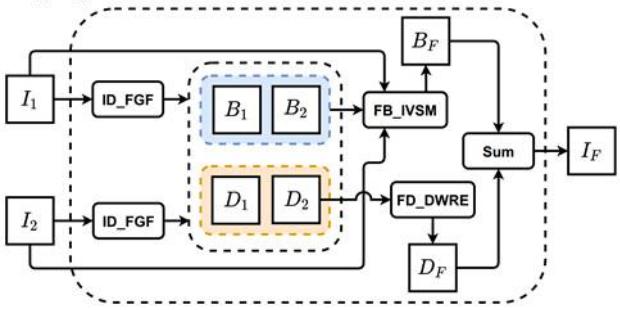


Figure 1. The proposed model

- **Step 1:** Use the IE_FGF method to decompose the two input images, I_1 and I_2, into base components B_1, B_2 and detail components D_1, D_2.

- **Step 2:** The base components B_1 and B_2 are fused using the FB_IVSM method to obtain the base component B_F.

- **Step 3:** The detail components D_1 and D_2 are fused using the FD_DWRE method to obtain the detail component D_F.

- **Step 4:** The final fused image I_F is obtained by summing the two components B_F and D_F.

4. Experimental results

4.1. Evaluation indicators

To evaluate the effectiveness of the proposed method, we use the following four evaluation metrics: Peak signal-to-noise ratio (PSNR), the sum of the correlations of differences (SCD), Entropy (E), Information preservation index in the fused image (Qabf).

4.2. Experimental setup

The experimental dataset consists of 42 pairs of infrared and visible images obtained from the TNO[1]

[1] https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029.

source. To evaluate the effectiveness of the proposed algorithm, we conduct comparisons using a variety of image fusion algorithms, including: DenseFuse (H. Li & Wu, 2019), FusionGAN (Ma, Yu, et al., 2019), PIAFusion (Tang et al., 2022), SDNet (Zhang & Ma, 2021).

4.3. Experimental results

The experimental outcomes are depicted in Figure 2 and detailed in Table 1. Visually, it is evident that the proposed method yields superior-quality images compared to alternative methods. Quantitatively, the proposed method outperforms all other methods across key evaluation metrics. It achieves the highest PSNR (63.4158), indicating superior image quality with minimal noise. In Entropy (6.8923) and Qabf (0.5608), it also ranks highest, demonstrating the best information preservation in the fused image. While DenseFuse has a slightly higher SCD (1.7838) than proposed method (1.7627), the difference is marginal, and proposed method still maintains strong correlation preservation. Compared to FusionGAN, which shows the lowest performance in multiple metrics, proposed method significantly improves fusion quality. Overall, the results confirm that proposed method produces the most balanced and informative fused images.



Figure 2. The fused image is generated by different image fusion algorithms

Table 1. Evaluation indexes obtained from fused images by image fusion algorithms

Dataset	Num	Algs	PSNR	SCD	E	Qabf
TNO	1	DenseFuse	62.5774	**1.7838**	6.8193	0.4457
	2	FusionGAN	60.9794	1.3793	6.5580	0.2341
	3	PIAFusion	61.7759	1.6032	6.8143	0.5281
	4	SDNet	62.1501	1.5590	6.6948	0.4294
	5	**Ours**	**63.4158**	1.7627	**6.8923**	**0.5608**

5. Conclusion and future work

In this paper, we proposed an improved infrared and visible image fusion method based on the Fast Guided Filter (FGF) and the IVSM. By leveraging FGF for image decomposition and IVSM for base component fusion, along with Distance-Weighted Regional Energy (DWRE) for detail component fusion, our method effectively preserves critical image features while enhancing contrast and structural details. Experimental results demonstrated that our approach outperforms existing methods in both subjective and objective evaluations, achieving superior results in terms of PSNR, entropy, and information preservation index.

In future work, we plan to explore optimization techniques to enhance processing efficiency while maintaining fusion quality. Additionally, we aim to investigate deep learning-based strategies to further refine the fusion process, leveraging convolutional neural networks (CNNs) or transformer-based architectures. Finally, expanding the evaluation to larger and more diverse datasets will help validate the robustness and generalizability of our method across different imaging conditions.

Acknowledgments

This research is funded by Thuyloi University Foundation for Science and Technology under grant number TLU.STF.24-01.

References

He, K., & Microsoft, J. S. (2015). *Fast Guided Filter*. https://arxiv.org/abs/1505.00996v1

Li, H., & Wu, X. J. (2019). *DenseFuse: A fusion approach to infrared and visible images.* IEEE Transactions on Image Processing, 28(5), 2614–2623. https://doi.org/10.1109/TIP.2018.2887342

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

Ma, J., Ma, Y., & Li, C. (2019). *Infrared and visible image fusion methods and applications: A survey.* Information Fusion, 45, 153–178. https://doi.org/10.1016/J.INFFUS.2018.02.004

Ma, J., Yu, W., Liang, P., Li, C., & Jiang, J. (2019). *FusionGAN: A generative adversarial network for infrared and visible image fusion.* Information Fusion, 48, 11–26. https://doi.org/10.1016/J.INFFUS.2018.09.004

Ma, J., Zhou, Z., Wang, B., & Zong, H. (2017). *Infrared and visible image fusion based on visual saliency map and weighted least square optimization.* Infrared Physics & Technology, 82, 8–17. https://doi.org/10.1016/J.INFRARED.2017.02.005

Meher, B., Agrawal, S., Panda, R., & Abraham, A. (2019). *A survey on region based image fusion methods.* Information Fusion, 48, 119–132. https://doi.org/10.1016/j.inffus.2018.07.010

Panigrahy, C., Seal, A., Gonzalo-Martín, C., Pathak, P., & Jalal, A. S. (2023). *Parameter adaptive unit-linking pulse coupled neural network based MRI–PET/SPECT image fusion.* Biomedical Signal Processing and Control, 83, 104659. https://doi.org/10.1016/J.BSPC.2023.104659

Tang, L., Yuan, J., Zhang, H., Jiang, X., & Ma, J. (2022). *PIAFusion: A progressive infrared and visible image fusion network based on illumination aware.* Information Fusion, 83–84, 79–92. https://doi.org/10.1016/J.INFFUS.2022.03.007

Yang, K., Xiang, W., Chen, Z., Zhang, J., & Liu, Y. (2024). *A review on infrared and visible image fusion algorithms based on neural networks.* Journal of Visual Communication and Image Representation, 101, 104179. https://doi.org/10.1016/J.JVCIR.2024.104179

Zhang, H., & Ma, J. (2021). SDNet: *A Versatile Squeeze-and-Decomposition Network for Real-Time Image Fusion.* International Journal of Computer Vision, 129(10), 2761–2785. https://doi.org/10.1007/S11263-021-01501-8/METRICS