

Optimized VLSI Implementation for Visible and Infrared Image Fusion using Stationary Wavelet Transform

Pitta Satya Surekha¹, Kandukuri Srinivas², Medipally Nagasri³

^{1,2,3}Assistant Professor, Department of ECE, Malla Reddy Engineering College and Management Sciences, Hyderabad, Telangana

Abstract

Image fusion has long relied on conventional signal processing techniques like discrete wavelet transform (DWT), contourlet transform, shift-invariant shearlet transform, and quaternion wavelet transform. However, these methods can introduce artifacts into the fused image, leading to suboptimal results. To address these issues, optimization-based fusion schemes have been proposed, although they often require multiple iterations to find the optimal solution, potentially resulting in oversmoothed images. This research focuses on a hardware-oriented VLSI-based implementation of visible-infrared (VI-IR) image fusion. The process begins by reading images in the MATLAB environment and applying stationary wavelet transform to decompose VI and IR images into multiple bands. Low-low bands are converted into text files, and a band fusion rule is applied using a multiplexer-based adder to combine both images. The final fused image is reconstructed in MATLAB. Simulation results demonstrate that this proposed method offers improved performance.

Keywords: Visible Image, Infrared Image, Stationary Wavelet Transform, Discrete Wavelet Transform, Image Fusion, VLSI Implementation.

1. Introduction

Visual information present in a scene can be captured as an image using a charge coupled device (CCD). The wavelength of the visible (VI) light that can be captured by CCD sensor ranges from 400 nm to 700 nm. However, in most of the computer vision applications, CCD or VI image alone is not sufficient to provide all the details of the scene. To extract more details, complementary images of the same scene should be captured by using multiple sensors of different modalities. This can be done by capturing images in wavelengths other than the VI band of electromagnetic spectrum. Usually, objects with more than 0 0 K emit radiation throughout the infrared (IR) spectrum. The energy emitted by these objects can be sensed by IR sensors and displayed as images for the end users. However, these images alone are not sufficient to provide an accurate description about the targeted scene. Hence, useful information of visual and infrared images must be combined in a single image for better understanding of a particular scene. The method or process of merging or combining various images in a single image for better perception of the scene is referred to as fusion. The main applications of image fusion are in digital photography, medical imaging, remote sensing, military, concealed weapon detection, night vision, target detection and tracking etc. Fusion has two broad categories called single sensor image fusion (SSIF) and multi sensor image fusion (MSIF). In SSIF, a single sensor is used to acquire several images of the same scene.

However, in MSIF multi modal sensors are used to capture complementary information. Digital photography is an example of SSIF whereas remaining applications such as medical imaging, remote sensing, military, concealed weapon detection, target detection and tracking come under MSIF. In this paper we focus on MSIF, on military, concealed weapon detection and night vision applications. In general, in all these applications VI and IR sensors are used to acquire complementary information of the targeted scene. In the military, VI sensor is used to acquire visible information such as battlefield. However, IR sensor is used to acquire vehicles, animals, soldiers, and camouflage information based

on temperature variations. In concealed weapon detection, VI sensor is used to acquire visible information whereas millimeter wave (MMW) sensor is used to capture hidden or concealed weapons. Similarly, in night vision as well IR and VI cameras are used to acquire complementary information. The research on VI and IR sensor image fusion is broadly divided into two main categories. They are spatial domain or single scale fusion techniques (SSF) and multi-scale fusion techniques (MSF). In SSF, the process of fusion is performed on the same scale as source images. However, MSF fusion process is performed on source images after decomposing them into multiple scales. MSF is further classified into three main categories. They are pyramid, wavelet and edge preserving decomposition-based methods.

Rest of the paper is organized as follows: Section 2 details about literature survey, section 3 details about the proposed methodology, section 4 details about the results with discussion, and section 5 concludes article with references.

2. Literature Survey

Jha, Ankit, Shirsha Bose, et al. (2023) [11] presented a novel fusion architecture called Global Attention based Fusion Network (GAF-Net), equipped with novel self and cross-attention learning techniques. They introduced the within-modality feature refinement module through global spectral-spatial attention learning using the query-key-value processing where both the global spatial and channel contexts were used to generate two channel attention masks. Since it was non-trivial to generate the cross-attention from within the fusion network, they proposed to leverage two auxiliary tasks of modality-specific classification to produce highly discriminative cross-attention masks. Vasu, G. Tirumala, et al. (2023) [12] presented a novel multi-focus image fusion method via local energy and sparse representation in the shearlet domain. The source images were decomposed into low- and high-frequency sub-bands according to the shearlet transform. The low-frequency sub-bands were fused by sparse representation, and the high-frequency sub-bands were fused by local energy. Bhalla, Kanika, et al (2022) [13] presented a multi-focus image fusion (MFIF) which integrated various input image features which produced single fused image using all its objects in focus. However, it is computationally complex, which led to inconsistency. Hence, the MFIF method was employed to generate the fused image by integrating the fuzzy sets (FS) and convolutional neural network (CNN) which detected focused and unfocused parts in both source images. It was also compared with other competing six MFIF methods like Neutrosophic set based stationary wavelet transform (NSWT), guided filters, CNN, ensemble CNN, image fusion-based CNN and deep regression pair learning (DRPL). Tang, Wei, et al. (2022) [14] presented an adaptive convolution for adaptively modulating the convolutional kernel based on the global complementary context. To further model long-range dependencies, an adaptive Transformer was employed which enhanced the global semantic extraction capability. Their network architecture was designed in a multiscale fashion so that useful multimodal information was adequately acquired from the perspective of different scales. Midde, Venkata Siva, et al. (2023) [15] presented an efficient image scaling algorithm and its architecture was developed which produced good quality resized image with lesser area and high performance. It used a linear space-variant edge detector for edge enhancement and a spatial sharp filter which reduced the blurring effects produced by the bilinear interpolation.

Xu, Xiaoli, Chenghua Fan, et al. (2022) [16] presented the QCA technology for image processing in various ways. The fundamentals of image processing approaches and the function of QCA in them were covered. The findings demonstrate that a more effective layout based on QCA technology reduced and computational complexity in the long run. Saranya, C., et al. (2022) [17] presented electronic form for superior storage and intelligent processing. This needed processing of documents

which utilized different processing strategies. Pre-processing techniques were advantageous in the recognition of document images as well as medical images. Meher, Bikash, et al. (2022) [18] presented a region-based image fusion technique using an efficient adaptive transition region extraction (ATRE) strategy. This method was beneficial because it did overcome the problems of noise sensitivity, poor contrast and blurring effects associated with the conventional pixel-based methods. The proposed ATRE technique was used to efficiently extract the bright object regions from the IR image and retain much of the visual background regions from the VI. An adaptive parameter was introduced for accurate segmentation. Han, Xiyu, et al. (2019) [19] proposed an adaptive two-scale image fusion method using latent low-rank representation (LatLRR). First, both IR and VI images were decomposed into a two-scale representation using LatLRR to generate low-rank parts (the global structure) and saliency parts (the local structure). The algorithm denoises at the same time. Then, the guided filter was used in the saliency parts to make full use of the spatial consistency, which reduced artifacts effectively. Panguluri, Sumanth Kumar, et al. (2021) [20] presented the Discrete wavelet transform based image fusion using unsharp masking. DWT was used for decomposing input images (infrared, visible). Approximation and detailed coefficients were generated. For improving contrast unsharp masking was applied on approximation coefficients. Then for merging approximation coefficients produced after unsharp masking average fusion rule was used.

3. Proposed Methodology

3.1 Overview of SWT

The fusion of visible Image (VI) and infrared (IR) images is an important task in many applications, such as surveillance, object recognition, and target tracking. Matlab is a widely used tool for image processing and image fusion, and it can be used in conjunction with VLSI design tools to implement image fusion algorithms in hardware.

In the context of VI and IR image fusion, Matlab can be used to develop and test image fusion algorithms, and to simulate their performance under different conditions. Once a suitable algorithm has been developed and tested, it can be implemented in hardware using VLSI design tools. There are several image fusion techniques that can be implemented using Matlab and VLSI design tools, including DWT-based fusion, fuzzy logic-based fusion, neural network-based fusion, and morphological operations-based fusion. These techniques can be optimized for low-power consumption and real-time processing using VLSI design tools, and the performance of the resulting hardware implementation can be evaluated using Matlab simulations.

Overall, the combination of Matlab and VLSI design tools offers a powerful platform for developing and implementing image fusion algorithms for VI and IR images. It allows for rapid prototyping and optimization of image fusion algorithms and can lead to efficient and high-performance hardware implementations. However, the complexity of the design process and the need for specialized expertise in both Matlab and VLSI design tools can be a limitation for some applications.

The DWT decomposes the input images into different frequency sub bands, which can be fused in a way that preserves important features from both images. Fuzzy logic-based fusion methods use membership functions to combine information from the input images, while neural network-based fusion can learn to identify important features and patterns from the input images. Morphological operations-based fusion can be used to enhance image contrast and reduce noise, which can improve the quality of the fused image.

Overall, the combination of Matlab and VLSI design tools offers a powerful platform for developing and implementing image fusion algorithms for VI and IR images. It allows for rapid prototyping and

optimization of image fusion algorithms and can lead to efficient and high-performance hardware implementations.

3.2 Proposed Model

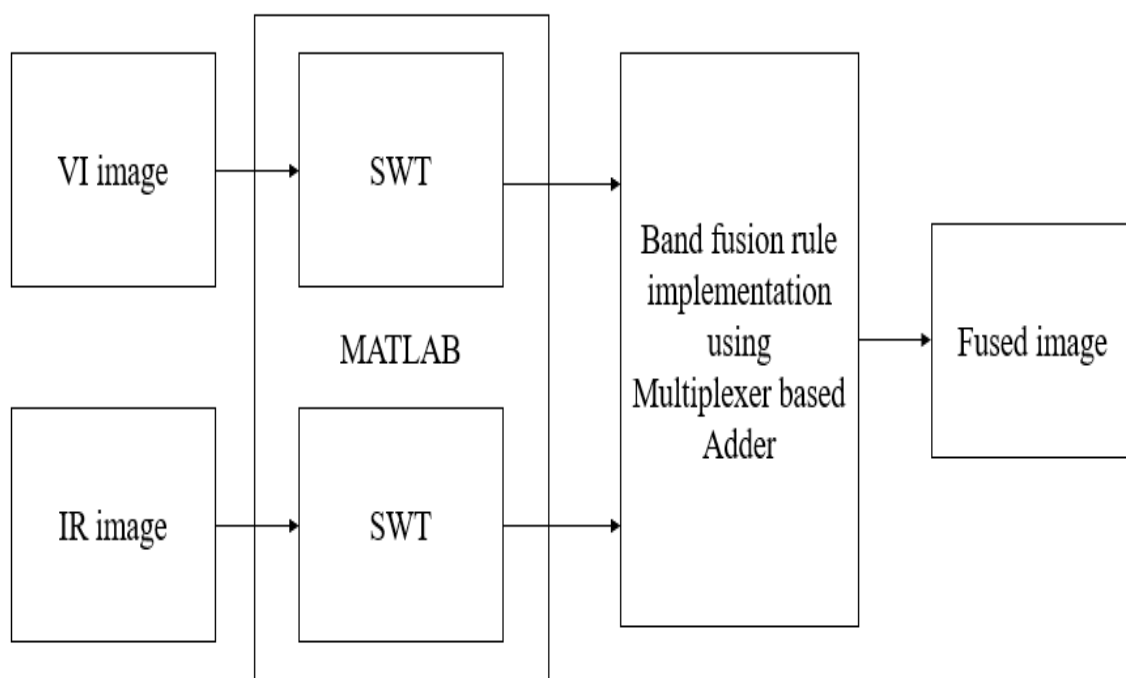


Figure 1. Proposed block diagram.

For decades, many conventional signal processing methods have been applied in the image fusion field to extract image features, such as discrete wavelet transform (DWT), contourlet transform, shift-invariant shearlet transform and quaternion wavelet transform etc. For the infrared and visible image fusion task. But these methods may introduce artifacts into the fused image. To overcome these problems optimization based fusion schemes are proposed. These methods take multiple iterations to find the optimal solution (fused image). These optimization methods may over smooth the fused image because of multiple iterations.

Figure 1 shows the proposed block diagram. This work focused on hardware-oriented implementation of VLSI based implementation of VI-IR image fusion mechanism. Initially, images are read using Matlab environment. Then, stationary wavelet transform is applied to decompose the VI, and IR images into multiple bands. The low-low bands are converted into text files. Further, the band fusion rule is applied on low-low bands, where multiplexer-based adder is used to combine both images. Finally, Matlab environment is used to reconstruct the fused image from the obtained text files. The simulation results shows that the proposed method resulted in better performance.

3.3 SWT

The stationary wavelet transform (SWT) is a signal processing technique that decomposes a signal into different frequency bands using wavelets. It is similar to the discrete wavelet transform (DWT), but unlike the DWT, the SWT does not downsample the signal at each level of decomposition, which leads to a shift-invariant representation of the signal. Here's a detailed algorithmic analysis of the stationary wavelet transform:

Step 1: Input: A signal $x(n)$ of length N , and a set of wavelet filters $h(n)$ and $g(n)$, where $h(n)$ is the low-pass filter and $g(n)$ is the high-pass filter.

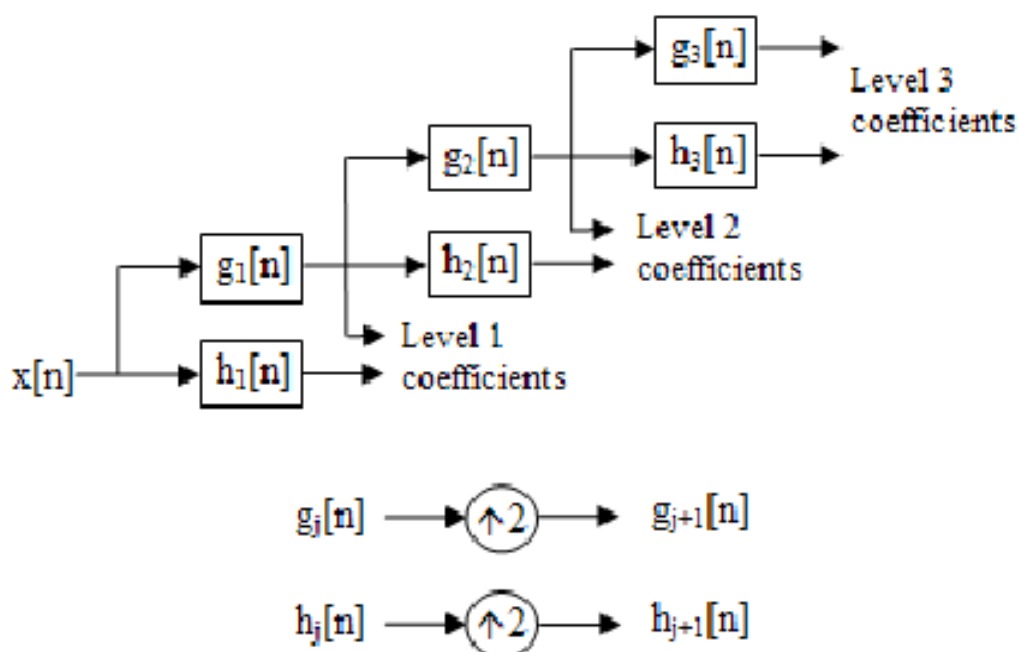


Figure 2. SWT

Step 2: Initialize a vector of approximation coefficients $V_0 = x(n)$.

Step 3: For each level of decomposition $i = 1$ to J , do the following:

a. Compute the wavelet coefficients $W_{j,i}(n)$ and the approximation coefficients $V_{j,i}(n)$ using the following equations:

$$W_{j,i}(n) = (V_{j-1,i} * g(n)) * 2^{i/2} \tag{1}$$

$$V_{j,i}(n) = (V_{j-1,i} * h(n)) * 2^{i/2} \tag{2}$$

where $V_{j-1,i}$ is the approximation coefficients from the previous level of decomposition.

b. Store the wavelet coefficients $W_{j,i}(n)$ and the approximation coefficients $V_{j,i}(n)$ in two separate matrices, W and V , respectively.

Step 4: Output: The wavelet coefficients W and the final approximation coefficients V_J .

Let's break down each step-in detail:

Step 1: Input The input to the SWT algorithm is a signal $x(n)$ of length N , and a set of wavelet filters $h(n)$ and $g(n)$, where $h(n)$ is the low-pass filter and $g(n)$ is the high-pass filter.

Step 2: Initialize We start by initializing a vector of approximation coefficients V_0 , which is just the original signal $x(n)$.

Step 3: Decomposition For each level of decomposition $i = 1$ to J , we compute the wavelet coefficients and approximation coefficients as follows:

- a. Compute the wavelet coefficients $W_{j,i}(n)$ and the approximation coefficients $V_{j,i}(n)$ using the following equations:

$$W_{j,i}(n) = (V_{j-1,i} * g(n)) * 2^{i/2} \quad (3)$$

$$V_{j,i}(n) = (V_{j-1,i} * h(n)) * 2^{i/2} \quad (4)$$

where $V_{j-1,i}$ is the approximation coefficients from the previous level of decomposition. These equations represent a convolution of the signal with the wavelet filters, followed by a decimation by a factor of $2^{i/2}$, where i is the level of decomposition. The decimation is not performed in the SWT, unlike in the DWT, which makes the SWT shift-invariant.

- b. Store the wavelet coefficients $W_{j,i}(n)$ and the approximation coefficients $V_{j,i}(n)$ in two separate matrices, W and V , respectively.

At each level of decomposition, we store the wavelet coefficients $W_{j,i}(n)$ and the approximation coefficients $V_{j,i}(n)$ in two separate matrices, W and V , respectively. These matrices will contain the coefficients for each level of decomposition.

Step 4: Output The final output of the SWT algorithm is the wavelet coefficients W and the final approximation coefficients VJ . The wavelet coefficients represent the high-frequency components of the signal at different scales, while the approximation coefficients represent the low-frequency components of the signal at the final level of decomposition.

3.4 Band fusion rule

The average band fusion rule is

$$LL_{new} = \{LL1 + LL2\} / 2$$

Here, addition operation between $LL1$, $LL2$ is performed using hybrid multiplexer-based adder.

A full adder is a combinational logic circuit that adds two binary digits (bits) and a carry input to produce a sum output and a carry output. One way to implement a full adder is by using an 8-to-1 multiplexer (mux). Here's a detailed analysis of a full adder using an 8-to-1 mux:

Step 1: Input: Two binary digits (bits) A and B , and a carry-in (C_{in}).

Step 2: Implementing half adders for the input bits A and B .

- Implement a XOR gate between A and B to generate a sum S_0 .
- Implement an AND gate between A and B to generate a carry-out C_0 .

Step 3: Implementing the 8-to-1 mux.

- The eight inputs to the mux are the carry-in C_{in} and the two bits A and B with all possible combinations of carry-in and bits ($C_{in}=0, A=0, B=0$), ($C_{in}=0, A=0, B=1$), ($C_{in}=0, A=1, B=0$), ($C_{in}=0, A=1, B=1$), ($C_{in}=1, A=0, B=0$), ($C_{in}=1, A=0, B=1$), ($C_{in}=1, A=1, B=0$), ($C_{in}=1, A=1, B=1$).
- The three select lines of the mux are S_0 , C_{in} , and their complement.
- The output of the mux is the sum S_1 and the carry-out C_1 .

Step 4: Implementing the final sum and carry-out.

- Implement a XOR gate between S_0 and S_1 to generate the final sum output S .

b. Implement an OR gate between C0 and C1 to generate the final carry-out output Cout.

Step 5: Output: The sum output S and the carry-out output Cout.

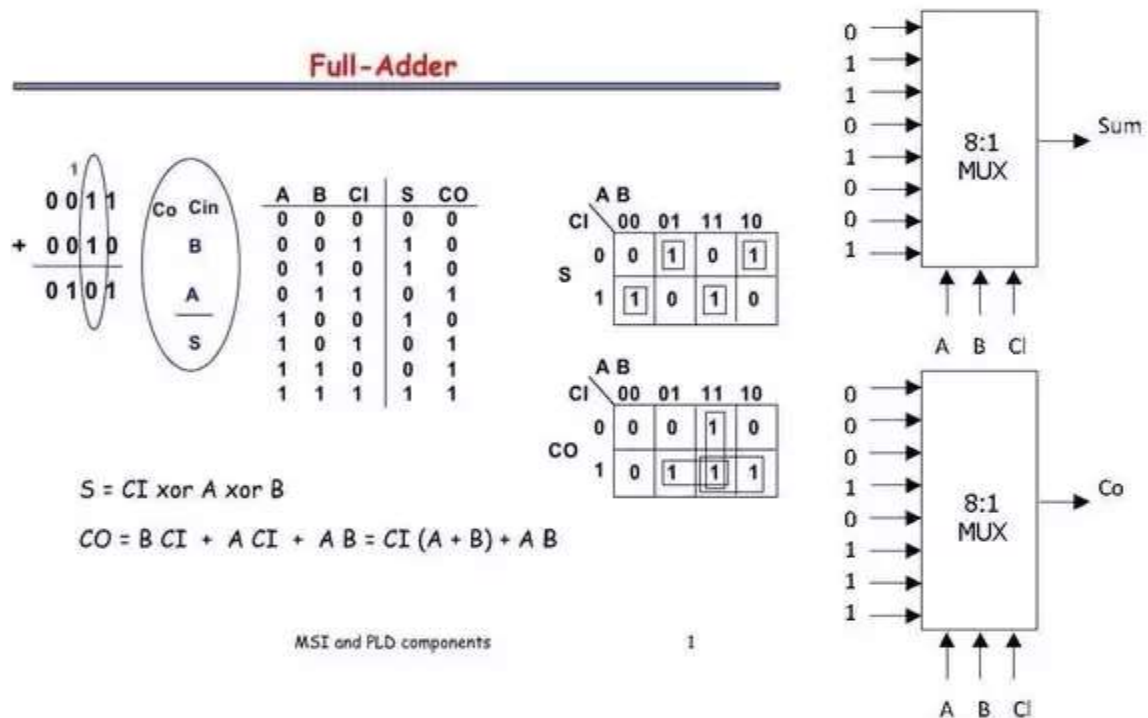


Figure 3. Full adder.

Hybrid adder:

A hybrid adder is a type of adder that combines multiple adder circuits to improve the speed and efficiency of the addition process. In this analysis, we will discuss the implementation of a hybrid adder using the above full adder that is implemented using an 8-to-1 multiplexer.

A hybrid adder can be constructed by dividing the binary number to be added into multiple sections, each of which can be added in parallel. The partial sums generated by each section are then added together using a final adder circuit to obtain the final sum.

To implement a hybrid adder using the full adder circuit implemented using an 8-to-1 multiplexer, we can use the following steps:

Step 1: Divide the binary number to be added into multiple sections, each containing n bits. Let's call them A0, B0, A1, B1, A2, B2, ..., An-1, Bn-1.

Step 2: Implement n instances of the full adder circuit using an 8-to-1 multiplexer, one for each pair of bits in the input sections. For example, one full adder circuit can be used to add A0 and B0, another to add A1 and B1, and so on.

Step 3: Connect the carry-out (C-out) output of the first full adder circuit to the carry-in (C-in) input of the second full adder circuit, and so on for all full adder circuits except for the last one.

Step 4: The output of the last full adder circuit will be the final sum of the input sections.

Step 5: If the binary number has an odd number of bits, one extra bit can be added to one of the sections to make the number of bits even, and the final carry-out bit from the last full adder circuit can be added to obtain the final sum.

The hybrid adder can provide faster performance compared to a single full adder circuit by performing addition in parallel. The time taken to obtain the final sum depends on the number of sections and the time taken by the final adder circuit to add the partial sums together. By carefully choosing the number of sections and the type of final adder circuit, the speed and efficiency of the addition process can be optimized.

4. Results and Discussions

All the experiments have been done in MATLAB 2016b version under the high-speed CPU conditions for faster running time. The aim of any fusion algorithm is to integrate required information from both source images in the output image. Fused image cannot be judged exclusively by seeing the output image or by measuring fusion metrics. It should be judged qualitatively using visual display and quantitatively using fusion metrics.



Figure 4. Fusion performance on Dataset-1

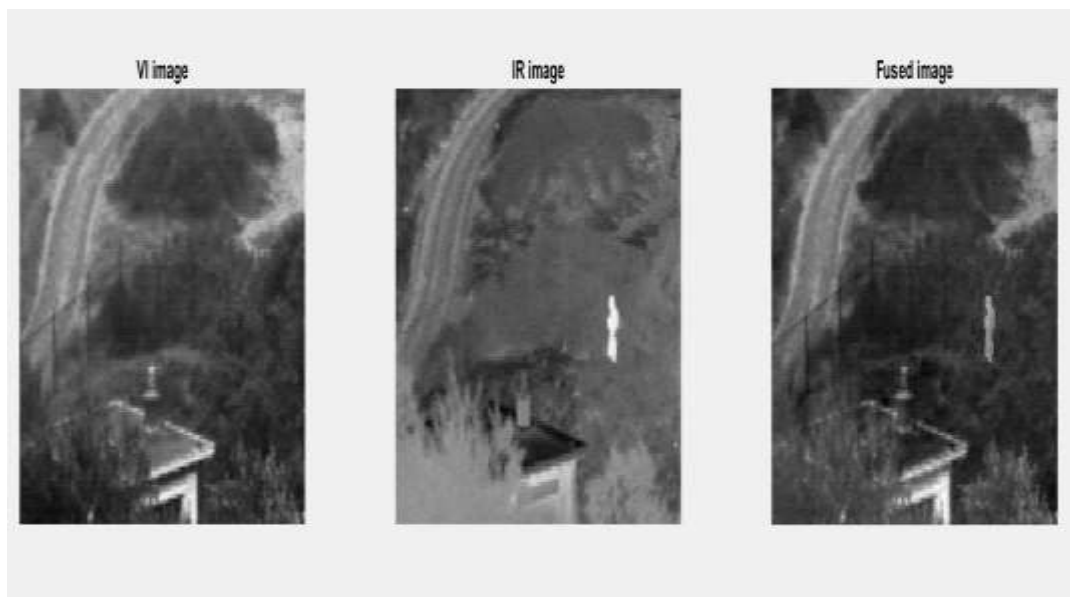


Figure 5. Fusion performance on Dataset-2

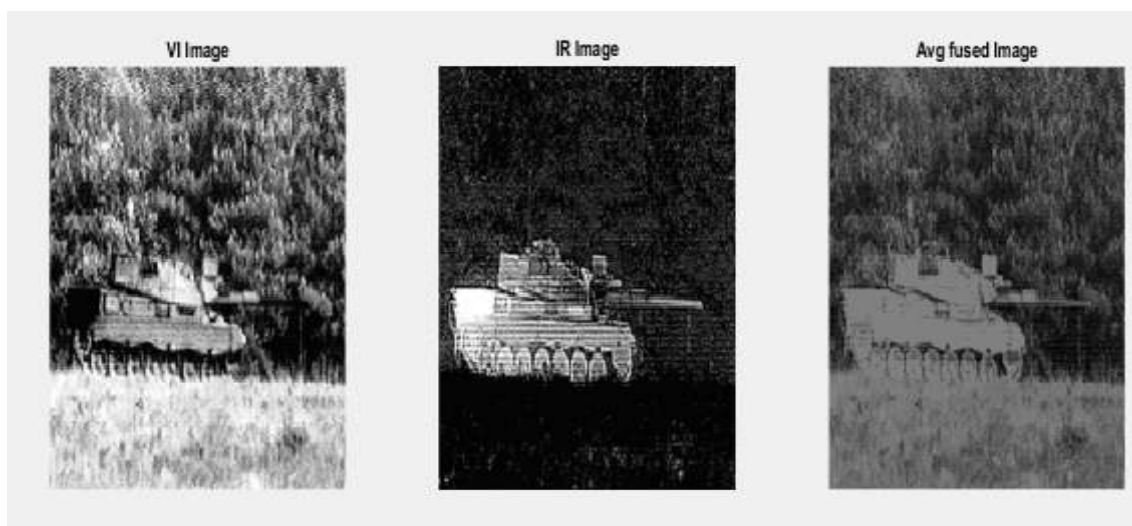


Figure 6. Fusion performance on Dataset-3



Figure 7. Fusion performance on Dataset-4

MATLAB RESULTS

Table 1 - Quantitative analysis of fusion methods for dataset in MATLAB

Methodology	PSNR (in dB)	RMSE	CC	SSIM	Entropy
DWT [25]	68.98	0.0906	0.934	0.988	0.9683
Proposed method	74.18	0.049	0.973	0.999	5.16

Quantitative analysis with IQA shown in table 1 for the test results presented in figure 4, which gives the analysis of dataset 1. Table 1 consists of various fusion metric parameters such as PSNR, RMSE, CC, SSIM and entropy. The best values are highlighted in bold letters. Our proposed method obtained far better values over all the existing fusion methods discussed in the literature. We also tested the qualitative analysis of dataset 2 with the similar fusion metric parameters considered for dataset.

VLSI RESULTS

Table 2. Quantitative analysis of fusion methods for dataset in VLSI

Parameters	Values
power	21.468w
Area (No of LUTS)	32
Delay	2.024

5. Conclusion

The hardware-oriented implementation of the VLSI-based implementation of the VI-IR image fusion mechanism was the primary focus of this work. In the beginning, the images are read in a Matlab

environment. The visible and infrared images are then subjected to a stationary wavelet transform so that they can be decomposed into multiple bands. Text files are created from the low-low bands after they are converted. In addition, the band fusion rule is applied to the low-low bands, and a multiplexer-based adder is utilised so that both images can be combined. In the final step, the Matlab environment is used to reconstruct the fused image from the text files that were previously obtained. The results of the simulation show that the method that was proposed ended up producing better performance.

References

- [1]. Hsia, Shih-Chang, Szu-Hong Wang, and Ting-Tseng Kuo. "VLSI architecture and implementation of HDR camera signal processor." *Journal of Real-Time Image Processing* 20.1 (2023).
- [2]. Jagtap, Nalini S., and Sudeep D. Thepade. "Improved Image Fusion Technique Using Convolutional Neural Networks and The Hybrid PCA-Guided Filter." *International Journal of Computing and Digital Systems* (2023).
- [3]. Chen, Shih-Lun, et al. "VLSI Design Based on Block Truncation Coding for Real-Time Color Image Compression for IoT." *Sensors* 23.3 (2023): 1573
- [4]. Mohankumar, M., et al. "A Novel Undistorted Image Fusion and DWT Based Compression Model with FPGA Implementation for Medical Applications." (2022).
- [5]. Rai, Ashruti, Vikrant Bhateja, and Anushka Dikshit. "Mammogram Image Fusion using Sigmoidal Function and Max. Fusion Rule." *Smart Intelligent Computing and Applications, Volume 2: Proceedings of Fifth International Conference on Smart Computing and Informatics (SCI 2021)*. Singapore: Springer Nature Singapore, 2022.
- [6]. Dikshit, Anushka, Vikrant Bhateja, and Ashruti Rai. "Discrete Wavelet Transform-Based Fusion of Mammogram Images for Contrast Improvement." *Advances in Micro-Electronics, Embedded Systems and IoT: Proceedings of Sixth International Conference on Microelectronics, Electromagnetics and Telecommunications (ICMEET 2021), Volume 1*. Singapore: Springer Nature Singapore, 2022.
- [7]. Vivek, C., Gayathri, P., Ranjitha, B., & Vibhu Harshini, M. (2022). VLSI Architecture for Image Scaling in Multimedia Applications. In *Soft Computing for Security Applications: Proceedings of ICSCS 2022* (pp. 423-433). Singapore: Springer Nature Singapore.
- [8]. Devanna, H., and K. Sudhakar. "The performance analysis of high-efficiency and low-power architecture for fuzzily-based image fusion has been carried out in this paper."
- [9]. Tang, Linfeng, et al. "DIVFusion: Darkness-free infrared and visible image fusion." *Information Fusion* 91 (2023): 477-493.
- [10]. Yue, Jun, et al. "Dif-Fusion: Towards High Color Fidelity in Infrared and Visible Image Fusion with Diffusion Models." *arXiv preprint arXiv:2301.08072* (2023).
- [11]. Jha, Ankit, Shirsha Bose, and Biplob Banerjee. "GAF-Net: Improving the Performance of Remote Sensing Image Fusion Using Novel Global Self and Cross Attention Learning." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2023.

-
- [12]. Vasu, G. Tirumala, and P. Palanisamy. "Gradient-based multi-focus image fusion using foreground and background pattern recognition with weighted anisotropic diffusion filter." *Signal, Image and Video Processing* (2023): 1-13.
- [13]. Bhalla, Kanika, et al. "A fuzzy convolutional neural network for enhancing multi-focus image fusion." *Journal of Visual Communication and Image Representation* 84 (2022): 103485.
- [14]. Tang, Wei, et al. "MATR: multimodal medical image fusion via multiscale adaptive transformer." *IEEE Transactions on Image Processing* 31 (2022): 5134-5149
- [15]. Midde, Venkata Siva, and E. P. Jayakumar. "Low-cost low-power approximated VLSI architecture for high-quality image scaling in mobile devices." *Journal of Real-Time Image Processing* 20.1 (2023): 11.
- [16]. Xu, Xiaoli, Chenghua Fan, and Lijun Wang. "A deep analysis of the image and video processing techniques using nanoscale quantum-dots cellular automata." *Optik* 260 (2022): 169036.
- [17]. Saranya, C., et al. "An Image Processing Application: Design of VLSI based Bilateral Filter using Distance Matrix Method." *2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*. IEEE, 2022.
- [18]. Meher, Bikash, et al. "Visible and infrared image fusion using an efficient adaptive transition region extraction technique." *Engineering Science and Technology, an International Journal* 29 (2022): 101037.
- [19]. Han, Xiyu, et al. "An adaptive two-scale image fusion of visible and infrared images." *IEEE Access* 7 (2019): 56341-56352.
- [20]. Panguluri, Sumanth Kumar, and Laavanya Mohan. "An enhanced image fusion framework using morphological operations based unsharp masking." *2021 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2021