Infrared-Visible Image Fusion in the Gradient Domain

by

Sanduni Premaratne
B.Sc., University of Moratuwa, 2016
M.Phil., University of Moratuwa, 2019

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF APPLIED SCIENCE

in the Department of Electrical and Computer Engineering

Infrared-Visible Image Fusion in the Gradient Domain

by

Sanduni Premaratne
B.Sc., University of Moratuwa, 2016
M.Phil., University of Moratuwa, 2019

Supervisory Committee

Dr. Panajotis Agathoklis, Co-Supervisor
(Department of Electrical and Computer Engineering)

Dr. Leonard T. Bruton, Co-Supervisor
(Department of Electrical and Computer Engineering)

Dr. Daniela Constantinescu, Non-Unit Member
(Department of Mechanical Engineering)

Abstract

Due to the complementary properties of the infrared cameras compared to conventional visible imaging cameras, it has become increasingly popular to fuse infrared and visible images of the same scene for better visual understanding. One major application of this is surveillance which involves videos and requires fast processing. Therefore, there is a need for investigating novel low-complexity fusion algorithms that can be implemented in real-time applications.

In this study, we address this critical research problem by two-scale fusion in the gradient domain with saliency detection and image enhancement. In the proposed method, the source images are first decomposed in to base and detail layers. Next, the base parts are fused in the gradient domain by choosing the maximum absolute gradient, whereas the gradients of the detail parts are fused using a weighted average where the weights are calculated using saliency maps. Prior to fusion, the detail parts are enhanced using a guided filter-based enhancement approach. Finally, the fused gradients of the base and detail components are added together to obtain the gradients of the fused image, from which the fused image is reconstructed using a reconstruction technique based on wavelets. Experimental results demonstrate that the proposed method achieves very competitive performance in subjective and objective fusion assessments, while also outperforming most methods in terms of computational complexity.

Index terms *- Image fusion, IR images, visible images, computational complexity, gradient domain*

Contents

List of Tables

List of Figures

List of Abbreviations

2D Two-Dimensional

3D Three-Dimensional

AG Average Gradient

CNN Convolutional Neural Network

CT Computed Tomography

DOF Depth Of Field

DWT Discrete Wavelet Transform

EI Edge Intensity

GAN Generative Adversarial Network

GPU Graphical Processing Unit

HDR High Dynamic Range

ICA Independent Component Analysis

IR Infrared

MEF Multi-Exposure Fusion

MFIF Multi-Focus Image Fusion

MMIF Multi-Modal Medical Image Fusion

MRI Magnetic Resonance Imaging

MSD Multi-Scale Decomposition

NIR Near-Infrared

NMF Non-Negative Matrix Factorization

NMI Normalized Mutual Information

NSCT Nonsubsampled Contourlet Transform

PCA Principal Component Analysis

PCNN Pulse-Coupled Neural Network

PET Positron Emission Tomography

$\mathbf{Q_{CV}}$ Chen-Varshney Metric

$\mathbf{Q_C}$ Cvejie's Metric

SD Standard Deviation

SF Spatial Frequency

SPECT Single Photon Emission Computed Tomography

SSIM Structural Similarity Index Measure

SVD Singular Value Decomposition

TE Tsallis Entropy

UAV Un-Manned Aerial Vehicle

VIF Visual Information Fidelity

ViT Vision Transformer

Acknowledgments

First, I would like to express my heartfelt gratitude to the co-supervisors Dr. Panajotis Agathoklis and Dr. Leonard T. Bruton for their mentorship, advice, inspiring discussions and patience. Next, I would like to thank Dr. Daniela Constantinescu for her advice as a member of my supervisory committee.

My gratitude also goes towards my M. Phil. advisor Dr. Chamira Edussooriya from the University of Moratuwa, for encouraging me to pursue this endeavor.

Next, I wish to thank my beloved husband Dinushan, my parents, and my sister for their unconditional love and support.

I am also grateful to Dr. Sarah Huber from the UVic Research Computing Services, whose guidance was invaluable in navigating the high performance computing systems of Digital Research Alliance of Canada.

Finally, I would like to thank Natural Sciences and Engineering Research Council of Canada and the University of Victoria for the for the financial support that helped me throughout this endeavor.

Chapter 1

Introduction

The goal of imaging is to relay information about a scene. Imaging sensors capture reflected energy of a scene or radiated energy of objects and converts it into an electrical signal, which will then be quantized to obtain a digital signal representation. These imaging sensors operate in different wavelength ranges of the electromagnetic spectrum and have contrasting utilization.

For computer vision, single-sensor imaging may be insufficient to capture the complete information about a target scene. It has been realized that sometimes we require two or more images of the same scene for better visual understanding. Depending on the application, this set of images can be captured using a single sensor at different times, or by multiple sensors of the same modality, or by cameras of different modalities. The images will then have complementary information about the same scene and thus require fusing them in to a single image to provide a fuller description than any of the individual images.

Depending on the application and the imaging modalities used, there are several categories in image fusion. They are, multi-exposure fusion (MEF), multi-focus image fusion (MFIF), multi-modal medical image fusion (MMIF), remote sensing image fusion, and infrared (IR)-visible image fusion. Figure 1.1 demonstrates some examples of these categories of image fusion.

In MEF, multiple images of the same scene are captured using different exposure levels. This is generally done by capturing images one after the other with changed exposure level. By fusing images taken at different exposures, MEF aims to create an image which resembles a high dynamic range (HDR) image [1].

In digital photography, some objects that are at different distances from the camera cannot be focused at the same time. The range of distances that can have an

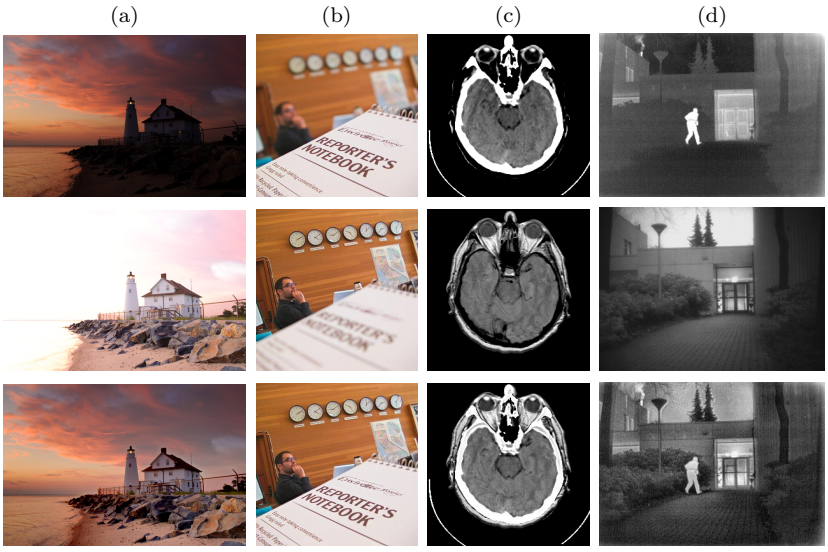


Figure 1.1: Categories of image fusion (a) MEF (b) MFIF (c) MMIF (d) IR-visible. First row depicts under-exposed image, near-focus image, CT image and IR image. Second row depicts over-exposed image, far-focus image, MRI image, visible image. Third row depicts the fused result of each image pair.

acceptable level of sharpness is denoted as the depth of field (DOF) and it depends on the focal length setting of the camera. For better visual clarity, sometimes we require that all objects in a scene are in focus. The approach to achieve this is MFIF, where we capture images at different focal lengths and combine them to generate an all-in-focus image [2].

For diagnosing diseases different medical imaging modalities have been used for a long time. Some well-known imaging modalities currently being used in clinical applications are, X-rays, computed tomography (CT), positron emission tomography (PET), single photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI). Often it is difficult to extract all the necessary information for a diagnosis from a single imaging modality. Therefore, MMIF aims to fuse images of the same site captured using different modalities thereby aiding the subsequent diagnosing process [3].

Currently, there are many earth-observation satellites such as SPOT, WorldView, IKONOS and Landsat, which monitor the geographical and environmental changes over time. These systems can obtain images using two different modalities; panchromatic and multi-spectral. The panchromatic images have high spatial resolution but low spectral resolution. In contrast to this, multi-spectral images have high spectral

resolution and low spatial resolution. The task of pan-sharpening aims to obtain fused images with both high spectral and spatial resolution [4, 5].

IR-visible image fusion aims to combine information from thermal imaging and visible imaging of the same scene. While visible imaging captures reflected light, infrared images capture the thermal radiation of objects and thereby helps to detect important targets in low-visibility areas. Fusion of these two modalities creates a more informative image which contains both texture characteristics from the visible image and target details from the IR image.

This thesis focuses on IR-visible image fusion. The main objective of the work presented here is to propose a novel method that has competitive quantitative and qualitative image fusion performance with lesser complexity, thereby achieving near real-time image fusion. The main contributions of the thesis are listed in the next section.

1.1 Contributions of the Thesis

The contributions in this thesis can be listed as follows.

- This thesis proposes a computationally efficient method for IR-visible image fusion. Here, a two-scale decomposition is employed and it is proposed to fuse base parts and detail parts separately in the *gradient domain* by utilizing two different fusion rules. For fusion of the base parts, maximum absolute gradient is chosen in order to increase information retention in the base layer. On the other hand, for detail layer fusion saliency maps are utilized. Here, the first convolutional layer of the pre-trained VGG19 network is utilized to obtain the saliency maps. These saliency maps will guide the fusion of the detail layers. In addition to this, a guided filter-based image enhancement scheme is employed to enhance the detail layers prior to fusion. Finally, the gradients of the fused base layer and fused detail layer are added together after which the image is reconstructed from the gradients.

- The performance of the proposed method is analyzed against 15 other methods in the literature. For this purpose, three aspects of performance evaluation, namely, quantitative fusion performance, qualitative fusion performance, and computational complexity, are considered.

- Experimental results demonstrate that the proposed method runs in ≈ 1 seconds on image pairs of size around 570×760, without hardware optimization or additional computational requirements such as graphical processing units (GPUs). Furthermore, the proposed method demonstrates significant quantitative fusion performance over 10 representative fusion metrics, as well as high qualitative performance.

1.2 Outline of the Thesis

The remaining chapters of this thesis are organized as follows.

- Chapter 2: Review on IR-Visible Image Fusion

  This chapter briefly describes different image fusion categories and elaborates on IR-visible image fusion in detail. Section 2.2.1 describes the different applications of IR-visible image fusion task whereas Section 2.2.2 elaborates on the types of currently available IR-visible image fusion methods.

- Chapter 3: IR-Visible Image Fusion in the Gradient Domain using Two-Scale Image Decomposition and Saliency Detection

  This chapter presents the proposed novel gradient domain IR-visible fusion algorithm based on two-scale decomposition and saliency detection. Section 3.2 reviews the algorithm utilized to reconstruct the fused image from the fused gradient field. Next, Section 3.3 reviews the guided filter which is utilized for enhancing the detail layers prior to fusion. Then, Section 3.4 presents the proposed fusion framework.

- Chapter 4: Experimental Results and Evaluation

  In this chapter, the experimental results are presented. Section 4.2 analyzes the quantitative performance of the proposed method in comparison with 15 other state-of-the-art methods in the literature, over 10 representative objective fusion metrics. Section 4.3 analyzes the qualitative performance of the proposed method and Section 4.4 compares the computational complexity.

- Chapter 5: Conclusions and Future Work

  This chapter presents the concluding remarks and directions for future work.

Chapter 2

Review on IR-Visible Image Fusion

2.1 Introduction

In this chapter, the problem of IR-visible image fusion is introduced. Section 2.2 elaborates IR-visible image fusion in detail, including motivation, applications and currently available methods for IR-visible image fusion. Section 2.3 details how performance evaluation of image fusion is conducted. Finally, Section 2.4 summarizes the contents of this chapter.

2.2 IR-Visible Image Fusion

All objects with a temperature above absolute zero (i.e. 0 K) emit electromagnetic waves. The wavelength of this emitted radiation depends on the temperature of the object. For most objects and environments, this thermal radiation is in the long-wave and mid-wave IR ranges. The thermal imaging cameras sense this radiation whereas active IR cameras capture reflected IR radiation by illuminating a scene with near-infrared (NIR) light [6].

On the other hand, visible imaging devices (i.e. cameras used for digital photography) capture the visible light reflected off the surfaces of objects in a scene. The visible spectrum consists of a narrow band of wavelengths from 400 nm to 700 nm whereas the IR range spans from 700 nm to 1 mm in the electromagnetic spectrum. Within the IR region, NIR, mid-wave, and long-wave regions are $700 - 1400$ nm, $3 - 8$ μm, and $8 - 15$ μm respectively [6].

Sensors operating in each of these IR spectral ranges provide different informa-

tion. For example, since passive IR imaging captures the thermal radiation emitted by objects, they provide information about the foreground such as target locations, vehicle/animal movements, weapon locations etc. However, they typically have low resolution and lacks texture information. One significant advantage of IR images captured in many of the aforementioned IR sub-bands is that they are resistant to disturbances such as poor illumination, fog and other adverse environmental effects.

In contrast to this, visible images are highly affected by adverse environment conditions that lead to poor visibility, since it affects the process of capturing reflected visible light. On the other hand, visible images have higher spatial resolution and provide information about the background with better texture details.

Thus, it is evident that IR and visible imaging have complementary properties which make the fusion of these two modalities highly beneficial compared to the use of either individual image. Therefore, IR and visible image fusion has become an important application for enhancing situational awareness in low visibility environments.

2.2.1 IR-Visible Image Fusion Applications

Typical applications of IR-visible image fusion include surveillance, object recognition, detection, image enhancement and remote sensing [7].

Object recognition

The goal of recognition systems is to specify the category of an object of interest. A highly prevalent recognition system is facial recognition which employs IR-visible image fusion. Omri et al. solves the problem of near-IR and visible image fusion for improving long range face recognition in [8]. This method utilizes wavelet transform to decompose the source images and applies singular value decomposition (SVD) or principal component analysis (PCA) for the fusion process. Hariharan et al. proposed a similar method using intrinsic mode functions for image decomposition instead of the wavelet transform [9].

Other face recognition methods using IR-visible image fusion include [10,11]. Heo et al. suggested adaptive fusion with eyeglass removal [10], whereas Bebis et al. proposed a wavelet domain-based method and a PCA-based method [11].

Detection and tracking

Detection based on IR-visible image fusion is widely found in many detection applications such as general object detection, fruit detection and pedestrian detection. Unlike recognition, detection and tracking applications require determining the position of the target object as well. In [12], He et al. proposed a target detection method using multi-level fusion and image enhancement. Other detection frameworks utilizing IR-visible image fusion include [13, 14]. Here, [13] focused on human detection whereas [14] targets moving object detection.

Target tracking generally involves videos as opposed to detection. In addition to doing object detection, tracking methods need to determine the relationship between video frames to map the target movement. Therefore, these methods need to be highly computationally efficient, so that they can track targets in real-time. Some tracking methods which utilize IR-visible image fusion can be found in [15, 16]. The work by Schenelle and Chan in [15] explored several fusion schemes for target tracking such as pixel averaging, PCA-based fusion, laplacian pyramid-based fusion, and wavelet decomposition-based fusion. Kumar et al. proposed using Haar wavelet decomposition and Fuzzy inference to fuse the source images for target tracking [16].

Surveillance

Similar to tracking, surveillance applications also typically deal with videos and require real-time implementations. Niu et al. used IR-visible fusion to improve the visibility of un-manned aerial vehicles (UAVs). Bhatnagar and Liu proposed a fusion scheme using nonsubsampled contourlet transform (NSCT) [17], whereas Paramanandham and Rajendiran proposed a discrete wavelet transform (DWT)-based fusion method for surveillance [18]. Tao et al. proposed an IR-visible image fusion method for space debris on-orbit surveillance [19].

Remote sensing

Remote sensing utilized IR-visible image fusion under various conditions. Chang et al. proposed an adaptive method for fusing visible and IR remote sensing images using multi-contourlet transform [20]. A similar method to fuse astronomical images was proposed in [21] using undecimated dual-tree complex wavelet transform instead of multi-contourlet transform. Other remote sensing applications include satellite image fusion [22] and urban object detection [23].

2.2.2 IR-Visible Image Fusion Methods

The currently available methods for IR and visible image fusion can be largely categorized in to the following groups, based on their methodology [7].

- Multi-scale decomposition (MSD)-based methods
- Sparse representation based methods
- Subspace-based methods
- Neural network/Deep Learning-based methods
- Saliency-based methods
- Hybrid methods
- Other methods

MSD-based methods

A vast majority of IR and visible image fusion methods in the literature utilize some form of MSD. These decompose the images into components of different scales. In general, MSD-based fusion comprises the following steps; (1) decomposing each source image in to a series of multi-scale components, (2) fusing the multi-scale components according to selected fusion rules (3) using inverse transform to reconstruct the fused image from the fused components. Selection of the transform for MSD and choosing the fusion rules are the two key aspects of these fusion methods.

The transforms used for MSD in IR and visible image fusion can be divided in to 5 categories, namely, (1) pyramid transform, (2) wavelet transform, (3) NSCT, (4) edge-preserving filters, (5) other MSD methods [7] . The pyramid transform proposed by Burt and Adelson in [24] suggested separating images in to different sub-bands of spatial frequency, in a pyramid structure. For IR and visible image fusion, Laplacian pyramid [25, 26], steerable pyramid [27, 28] and contrast pyramids [29, 30] have been commonly used.

The wavelet transform is a concept widely used in the fields of image and signal processing. Typical wavelet transforms commonly utilized in IR and visible image fusion include DWT [31–33] and dual-tree DWT [34, 35]. The NSCT is a shift-invariant model that was proposed to overcome the issue of discontinuities when

extending 1-D wavelet transform to two dimensions. NSCT has been widely used in IR and visible image fusion due to its shift-invariance property which also reduces the effect of registration errors [36–39].

Basic smoothing filters and edge-preserving filters have also been increasingly used to decompose images in to one or more detail layers and a smooth base layer [40]. Some of the most commonly used filters include mean filter [41], bilateral filter [42,43], weighted least squares filter [44, 45] and guided filter [46, 47].

While MSD-based methods are simple and faster compared to many other methods, the number of decomposition levels needs to be increased for higher fusion quality, which in turn increases the computational complexity.

Sparse representation based methods

These methods aim to learn an over-complete dictionary using numerous high-quality natural images. Then, this learned dictionary is used to represent the source images. The sparse representation-based fusion methods generally consist of the following steps (1) decomposing each source image in to overlapping patches using a sliding window strategy, (2) learning an over-complete dictionary from a large selection of natural images (3) performing sparse coding on the patches obtained in the first step, (4) fusing the sparse representation coefficients according to a selected fusion rule, and finally (5) reconstructing the fused image from the fused coefficients using the same over-complete dictionary. Some methods that fall in to this category could be found in [48–50].

Since these methods divide source images in to overlapping patches, they display a robustness to registration errors [51]. However, these methods are generally slow, since sparse coding is a time-consuming process.

Subspace-based methods

The subspace-based methods project high-dimensional input images into lower-dimensional subspaces. Some popular sub-space methods are PCA and independent component analysis (ICA) [52].

PCA converts possibly correlated variables into uncorrelated variables, which are called principal components. This process reduces the redundant information and highlights important features [53]. Therefore, many IR-visible fusion methods employing PCA have been proposed to date [54–59].

In contrast to PCA, the projection to the sub-space found via ICA is both uncorrelated and independent [52]. Some fusion methods utilizing ICA can be found in [60–63].

Another dimensionality reduction method that has been used for IR-visible image fusion is non-negative matrix factorization (NMF) [64, 65].

Neural networks/Deep Learning based methods

In the past, most neural network-based IR and visible fusion methods were based on pulse-coupled neural networks (PCNNs). However, in the last few years, with the increase of IR and visible image data availability, shallow convolutional neural network (CNN) [66], deep CNN [67–69], generative adversarial network (GAN) [70] or convolutional encoder-decoder [71,72] architectures have been frequently proposed. Also, with the recent rise of the vision transformer (ViT) [73], currently there is a trend towards transformer-based architectures [74–76]. A comprehensive review of deep learning-based IR and visible image fusion techniques can be found in [77, 78].

While these methods demonstrate significant improvement over the traditional approaches, there are some drawbacks as well. Vast majority of the newer deep learning-based methods are focused on achieving superior fused image quality and therefore, these models are rather deep and complex. Thus, they require considerable processing power including GPUs and high memory for training. Furthermore, from the evaluation of run-time performance, it is evident that available models require a GPU to run fusion in real-time.

Saliency-based methods

Image saliency detection aims to find which objects and features attract human visual attention. Saliency-based fusion aims to improve the fused image quality by maintaining the integrity of the salient regions. For IR-visible image fusion saliency has been adopted in two ways: fusion weight calculation and salient object extraction.

In the first approach, saliency detection is adopted for reconstructing the fused image. A saliency detection method is utilized to obtain saliency maps for each source image, and then the fusion weight maps are obtained from those saliency maps. Some examples utilizing this approach are [79–82].

In the second approach, a saliency model is used to extract the regions of importance which is then used to guide the fusion, rather than directly computing weights

from the saliency map [83–85].

Hybrid methods

Hybrid methods are those that aim to integrate the advantages of two approaches by combining them. Hybrid of MSD and sparse representation [51, 86], hybrid MSD and neural network [87, 88], hybrid MSD with two decomposition branches [89, 90] are common hybrid approaches available in the literature.

Since the basis of hybrid methods is to combine multiple approaches, this in turn increases their computational complexity.

Other methods

A series of methods [91, 92] were proposed by Ma et al. using the total variation framework, which aim to simultaneously retain the intensity information from the IR image and the texture detail from the visible image. Guo et al. [93] further improved this approach by retaining intensity from both IR and visible images, while simultaneously preserving the texture information from the visible image.

2.3 Performance Evaluation of Image Fusion

As discussed above, IR-visible fusion has been adopted in different applications such as surveillance, object detection and tracking. The performance of these applications which utilize IR-visible image fusion, highly depend on the quality of the fused image. Thus, it is necessary to perform qualitative and quantitative evaluation on different fusion methods.

Traditionally, the quality of a fused image was evaluated by human observers. This subjective evaluation is a reliable assessment based on the human visual system. Thus, it can represent human expectations better. However, it also has the drawbacks of human bias, high cost, considerable time consumption, and the results not being reproducible [94]. Therefore, objective evaluation methods have been increasingly proposed and adopted in image fusion literature.

In contrast to subjective methods, objective evaluation methods produce a quantitative measure of the fused image quality and are not influenced by human bias. The metrics that are currently being used for objective evaluation can be grouped in to

four categories based on how the metric is formulated. They are, information theory-based metrics, structural similarity-based metrics, image feature-based metrics, and human perception-inspired metrics [95]. It is also important to note that most of the recently proposed fusion metrics attempt to emulate the human perception of image quality. In Chapter 4, 10 objective fusion metrics are utilized to analyze the performance of the proposed approach with respect to that of the other methods in the literature. The objective metrics used for this analysis are presented in Appendix A.

Since applications such as surveillance and target tracking deal with videos, it is imperative that IR-visible image fusion methods employed for such tasks are fast implementations. Therefore, the computational complexity of each fusion method also needs to be evaluated. In image fusion literature, this is generally done by calculating the average time taken to fuse IR-visible image pairs in an evaluation dataset.

2.4 Summary

In this chapter, we briefly described the image fusion research area, and presented a discussion on IR-visible image fusion with more detail. There, we discussed why fusing IR images and visible images of the same scene is desired. Furthermore, we elaborated on applications and currently available methods for IR-visible image fusion. Finally, we briefly described how performance evaluation of image fusion is conducted. In the next chapter, we present a novel method which aims to address some gaps in IR-visible image fusion literature.

Chapter 3

IR-Visible Image Fusion in the Gradient Domain using Two-Scale Image Decomposition and Saliency Detection

3.1 Introduction

In this chapter, a novel IR-visible image fusion method is proposed by utilizing multi-scale image decomposition and gradient-domain fusion.

Fusion in the gradient domain has had significant success in MEF and MFIF [96, 97] leading to high quality fusion results with lesser computational complexity. The approach in [96] is based on a method for image reconstruction from gradients using wavelets developed for adaptive optics applications [98]. So far, gradient domain-based approaches have not been considered for IR and visible image fusion. On the other hand, a vast majority of IR and visible image fusion methods in the literature utilize some form of MSD. The main advantage of MSD-based methods is that the decomposition allows one to treat different spatial frequency bands separately. Some key approaches used for decomposition are, pyramid transform [25, 28, 30], wavelet transform [31–33], and edge-preserving filters [43, 45–47, 99]. While these MSD-based methods have demonstrated good fusion performances, they also can cause low contrast, loss of detail, and artifacts. Furthermore, most MSD-based methods require more than two decomposition levels to obtain satisfactory fusion performance. This

in turn increases the computational complexity of MSD-based methods.

To combine the advantages of image fusion in the gradient domain and MSD, we propose a novel gradient domain fusion method using two-scale image decomposition and saliency detection. In the proposed method, the source images are first decomposed in to base and detail layers. Next, the base parts are fused in the gradient domain by choosing the maximum absolute gradient, whereas the detail parts are fused using a weighted average where the weights are calculated using saliency maps. Finally, the fused gradients of base and detail components are added together to obtain the gradients of the fused image, from which the fused image is reconstructed.

The subsequent sections of this chapter are dedicated to elaborating the proposed method. To this end, we first review the algorithm for reconstructing an image from its gradient field in Section 3.2. Section 3.3 reviews the guided filter which is utilized within the proposed framework for two-scale decomposition and image enhancement. Next, the proposed method is detailed in Section 3.4. Finally, the chapter is concluded with a summary in Section 3.5.

3.2 Image reconstruction from a gradient field : A Review

The problem of image reconstruction from the gradient data can be formulated as follows: given a set of image gradients $\Phi = [\Phi^x, \Phi^y]^T$, find I such that,

$$\nabla I = \Phi \tag{3.1}$$

where $\nabla = [\partial/\partial x, \partial/\partial y]^T$. Equation (3.1) will have a meaningful solution if the gradient Φ is a conservative vector field. For two-dimensional (2D) functions such as images, this is equivalent to the zero curl condition given by,

$$\frac{\partial^2 \Phi}{\partial x \partial y} = \frac{\partial^2 \Phi}{\partial y \partial x}. \tag{3.2}$$

A common approach is to formulate (3.1) as an optimization problem [100], which yields the Poisson equation,

$$\nabla^2 I = \nabla^T \cdot \Phi \tag{3.3}$$

where $\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$ denotes the Laplacian of I, and symbol $\cdot$ denotes the standard inner product [101].

The Poisson equation in (3.3), can be solved using different approaches. In the proposed method, we utilize the approach by Sevcenco et al. [101,102] which is based on Haar wavelets and inspired by [98]. The basic idea of this reconstruction method is the relationship between the Haar wavelet filters and the Hudgin gradient model [101].

The Hudgin gradients of an image I in x and y directions are given by [101],

$$\Phi^x(x,y) = I(x+1,y) - I(x,y) \tag{3.4}$$

$$\Phi^y(x,y) = I(x,y+1) - I(x,y). \tag{3.5}$$

The Haar analysis filters are given by [103],

$$H_L(z) = \frac{1}{\sqrt{2}}(1+z^{-1}) \tag{3.6}$$

$$H_H(z) = \frac{1}{\sqrt{2}}(1-z^{-1}). \tag{3.7}$$

Then, the relationship between the Hudgin gradient geometry and the Haar analysis filters can be specified as follows,

$$\Phi^x(x,y) = \sqrt{2}H_H(z_x)I \tag{3.8}$$

$$\Phi^y(x,y) = \sqrt{2}H_H(z_y)I \tag{3.9}$$

where $H_H(z_x)I$ and $H_H(z_y)I$ denote filtering I in x and y directions with $H_H(z)$. Here, z_x indicates horizontal filtering of rows whereas z_y indicates vertical filtering of columns.

The reconstruction technique from the gradients, is based on a two step process. In the first step the Haar wavelet decomposition coefficients of the image are obtained directly from the gradients. Detailed description of this analysis step can be found in [101]. Then, in the second step (synthesis), the image is reproduced from these Haar decomposition coefficients. Since gradient-domain image fusion is done by mixing gradients from multiple images, it is possible for the fused gradient field to violate the zero curl condition, thus violating (3.1). To avoid any artifacts in the reconstructed image, an iterative Poisson solver is utilized at each resolution level during the synthesis step. This iterative Poisson solver in the reconstruction algorithm

from [101] is a vital step for image fusion in the gradient domain.

The recursion formula can be given as [101],

$$I(k+1) = I(k) - \frac{1}{4}\left(\begin{bmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix} \otimes I(k) + \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \otimes \Phi^x(k) + \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \otimes \Phi^y(k)\right) \tag{3.10}$$

where k is the iteration index and $\otimes$ represents the 2D convolution. A very small number of iterations is sufficient since the initial point is good enough to give fast convergence. Number of iterations was chosen to be 3 in this paper. The drawback of the iterative Poisson solver is that it has a smoothing effect on the reconstructed images.

The advantage of this algorithm over other reconstruction methods is that it has a low complexity O(N) where N is the number of samples in the signal to be reconstructed. A detailed evaluation regarding the computational complexity can be found in [98].

3.3 Guided filtering : A Review

The guided filter is an edge-preserving smoothing filter proposed by He et al. and its computational complexity is independent of the filter size [104]. Furthermore, the guided filter is established on a local linear model, which makes it applicable for image processing applications such as image enhancement, colorization, and image matting [104]. In a later work [105], He et al. proposed a faster version of the guided filter.

The mathematical formulation of guided filtering assumes that the output image O is a linear transformation of the guidance image G within a local window ω_k centered at pixel k [104].

$$O_i = a_k G_i + b_k \qquad \forall i \in \omega_k \tag{3.11}$$

Within ω_k, the coefficients a_k and b_k are invariant. Therefore, they can be determined by minimizing the squared difference between the input image I and output image O

as follows.

$$E(a_k, b_k) = \sum_{i \in \omega_k} \left((a_k G_i + b_k - I_i)^2 + \epsilon a_k^2 \right) \tag{3.12}$$

Here, ϵ is a user-specified regularization parameter that has been introduced to prevent a_k from getting too large.

The coefficients a_k and b_k can be obtained using linear regression as follows.

$$a_k = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} G_i I_i - \mu_k \bar{I}_k}{\sigma_k + \epsilon} \tag{3.13}$$

$$b_k = \bar{I}_k - a_k \mu_k \tag{3.14}$$

Here, μ_k and σ_k are the mean and the variance of G in ω_k. $\bar{I}_k$ is the mean of I in ω_k and $|\omega|$ is the number of pixels in ω_k. Furthermore, $\frac{1}{|\omega|} \sum_{i \in \omega_k} G_i I_i - \mu_k \bar{I}_k$ term gives the covariance of G and I within ω_k.

When the linear model in (3.11) is applied to the entire image, each pixel will be contained in more than one window. To solve this problem of getting different values for the same pixel depending on ω_k location, all possible a_k and b_k values are averaged.

Then, the filter output is given by,

$$O_i = \bar{a}_i G_i + \bar{b}_i \tag{3.15}$$

where $\bar{a}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} a_k$ and $\bar{b}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} b_k$.

The biggest advantage of guided filtering is that its time complexity is $O(N)$ where N is the number of pixels. This is further improved by the fast guided filter [105]. In the fast guided filter, the input is downsampled prior to calculating $\bar{a}_i$ and $\bar{b}_i$, and the original size is recovered via upsampling. For a downsampling ratio of s, the time complexity becomes $O(N/s^2)$, since the number of pixels gets reduced to $O(N/s^2)$.

In the subsequent sections the fast guided filtering operation is denoted as $FGF(G, I, r, \epsilon, s)$, where G is the guidance image, I is the input image, r is the filter size, ϵ is the regularizing coefficient, and s is the downsampling ratio.

3.4 Proposed fusion framework

The proposed fusion method converts each image to the gradient domain, and fuses the gradients according to the selected fusion rules, after which the fused image is obtained via the inverse transformation. It is inspired by the MEF and MFIF method proposed by Paul et al. in [96]. This method uses simple maximum absolute rule for the fusion of the gradients. In contrast to this, the proposed method utilizes saliency maps for fusing the gradients of the detail components after a two-scale decomposition. As demonstrated by the experimental results, this leads to higher detail retention than [96] at no significant increase in run time. The schematic diagram of the proposed method is depicted in Figure 3.1.

The proposed method can be briefly summarized as follows. First, the image pair is decomposed with a two-scale decomposition using a edge-preserving smoothing filter operation. For this purpose, the fast guided filter [105] is utilized. The base components are then fused in the gradient domain using the maximum absolute gradient approach.

Next, the fusion weights for the detail components are calculated by obtaining the saliency maps using a pre-trained CNN. For this purpose, the VGG-19 model is used. For simplicity, only the first convolutional layer (conv1-1 in Figure 3.2) of VGG-19 model is considered. Then, the L_1 norm is calculated over the channels of the conv1-1 output to obtain the saliency maps. Finally, the weight maps are obtained by normalizing the saliency maps.

Then, the detail components are enhanced by utilizing a guided filter in order to alleviate the adverse smoothing effect of the Poisson solver. Using the weight maps obtained before, we take the weighted average of the gradients of the enhanced detail components. Then, the gradients of the base and detail components are summed together to obtain the gradients of the fused image. The fused image is reconstructed from the fused gradients using the reconstruction algorithm discussed in Section 3.2. Finally, the fused image is adjusted using histogram equalization with 8×8 tiles and 0.005 clip limit [106].

This method can be easily adapted for fusing IR-visible image pairs where the visible image is color. The color visible image is first converted to YCbCr and the luminance (Y) channel of the visible image is fused with the grayscale IR image. After this, we use histogram equalization on the fused luminance. The chrominance (Cb/Cr) channels of the visible image are taken as the chrominance of the fused

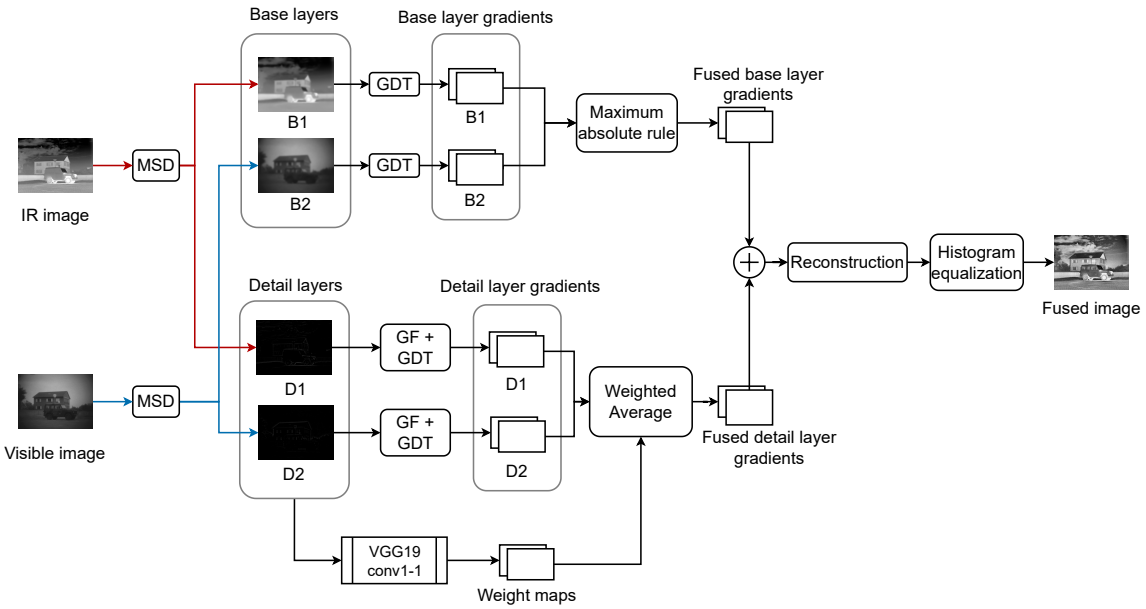


Figure 3.1: Schematic diagram of the proposed method. GDT denotes the transformation to the gradient domain, GF denotes guided filtering-based enhancement, and MSD denotes image decomposition.

image. Finally, the fused image is converted from YCbCr to RGB.

3.4.1 Two-scale image decomposition

MSD is a highly popular technique utilized in many IR and visible image fusion methods in the literature. The aim in MSD is to separate an image into different detail layers and base layers. Recently, edge-preserving smoothing filters are increasingly being used for this task. Among these, the guided filter [104] is a popular choice due to its linear time complexity. Utilizing the same approach, we propose using the fast guided filter [105] for a two-scale decomposition. The decomposition thus obtained can be expressed as,

$$B_n = FGF(G_n, I_n, r, \epsilon, s) \tag{3.16}$$

$$D_n = I_n - B_n. \tag{3.17}$$

where $n \in \{1, 2\}$ and I_1, I_2 are the IR image and the visible image respectively. For the fast guided filtering operation, we set $G_n = I_n, r = 4, \epsilon = 0.4^2, s = 4$ for both the source images. The parameter values for r, ϵ, and s were chosen empirically by testing various values and choosing those that demonstrated sufficient smoothing in

base layers.

The main advantage of image decomposition is that it facilitates using different fusion rules for different spatial frequency bands. By using a two-level decomposition, we aim to have a better trade-off of this advantage with a smaller increase in computational time.

3.4.2 Guided filtering-based image enhancement

As mentioned in Section 3.3, the guided filter is an edge-preserving filter. Thus, it can be utilized to enhance images as demonstrated in [104]. In IR-visible image fusion, Guo et al. [99] incorporated this technique for post-fusion image enhancement. In contrast to their method, the proposed method enhances the detail layers prior to fusing them in the gradient domain.

The fast guided filtering-based image enhancement process is employed on the detail layers as follows,

$$T_n = FGF(D_n, D_n, r, \epsilon, s) \tag{3.18}$$

$$O_n = (D_n - T_n)k + T_n \tag{3.19}$$

where k is the enhancement coefficient and T_n is the result of guided filtering D_n with itself as the guidance image. (3.19) extracts the most important edge information in D_n and then superimposes it on T_n, thus producing a detail enhanced image. For this enhancement step, empirically chosen values $r = 4, \epsilon = 0.1^2, s = 2, k = 2$ are used for both the source images. These values were chosen empirically by testing various values and choosing those that demonstrated best subjective fusion performance.

This step is intended to diminish the smoothing effect of the Poisson solver in the image reconstruction method. Also, it ensures that the resulting fused image has enhanced details than many comparative methods in the literature.

3.4.3 Fusion of detail layers

For fusing the detail components, we utilize saliency maps. A saliency map depicts the activity level or salient objects in an image. Image saliency detection is a currently highly popular research area in computer vision, and there is a plethora of methods proposed for this task. While simpler methods are available in literature, newer methods are generally based on deep learning.

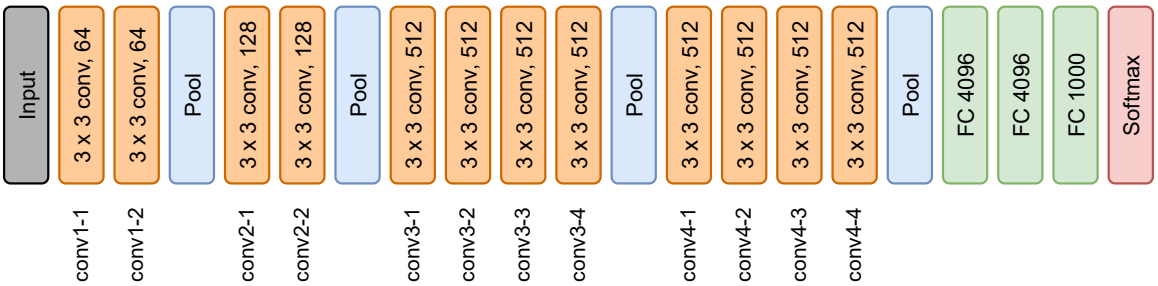


Figure 3.2: VGG19 architecture.

In the proposed method, we utilize a pre-trained CNN to generate the saliency maps. This approach has been successfully utilized to calculate saliency maps and fusion weights in IR and visible image fusion [107] and MEF. The key idea behind using pre-trained classification CNN models is that the early layers of these networks have learned the filters to extract the basic features of natural images. Thus, the output feature maps can provide a measure of the activity level of different parts of the images.

Unlike previous works such as DLF [107] where pre-trained CNNs were used for obtaining saliency maps, in the proposed method, only the conv1-1 layer (See Figure 3.2) of the VGG-19 network [108] is considered. The reasons for this choice are two-fold. (1) It was observed that utilizing multiple layers does not significantly improve the fusion performance but greatly increases the computational complexity. (2) Due to the pooling operations within the network, output from the deeper layers of these pre-trained CNNs are spatially quite smaller compared to the original image. Therefore, the saliency maps obtained from latter layers need to be upsampled to match the size of the original set of images. Due to this, the weight maps will contain upsampling artifacts which will worsen with larger upsampling ratios.

Let D_n denote the set of detail components corresponding to the pair of pre-registered source images I_n where I_1 is the IR image and I_2 is the visible image. Additionally, consider the first convolutional block of pre-trained a CNN model, output of which contains C output channels.

The C channel output feature map is given by,

$$f_n = \max(0, F_1(D_n)) \tag{3.20}$$

where $F_1(D_n)$ is the output from feeding D_n through the conv1-1 block. The function $\max(0, .)$ denotes the ReLU operation.

Next, the L_1 norm of f_n is obtained over its C channels as follows,

$$\hat{f}_n = \sum_{c=1}^{C} |f_n^c| \tag{3.21}$$

where f_n^c denotes the c-th channel of the output feature map f_n.

Then, $\hat{f}_n$ is taken as the saliency maps corresponding to the n-th input image since it gives a measure of the activity level. The saliency maps $\hat{f}_n$ are normalized to obtain the weight maps w_n.

$$w_n = \frac{\hat{f}_n}{\sum_{n\in\{1,2\}} \hat{f}_n} \tag{3.22}$$

Finally, the fused detail part gradients are obtained via,

$$\Phi_D^x = \sum_{n\in\{1,2\}} w_n \Phi_{D_n}^x \tag{3.23}$$

$$\Phi_D^y = \sum_{n\in\{1,2\}} w_n \Phi_{D_n}^y. \tag{3.24}$$

3.4.4 Fusion of base layers

As mentioned before, two different fusion rules are adopted for base components and detail components. For the fusion of the two base parts, we utilize the maximum absolute rule in the gradient domain.

Let I_n denote a set of pre-registered source images where $n \in \{1, 2, \ldots, N\}$. Also, let B_n denote the corresponding base components obtained via (3.16). According to the Hudgin gradient model introduced before, the gradients of these base components are given by,

$$\Phi_{B_n}^x(x, y) = B_n(x + 1, y) - B_n(x, y) \tag{3.25}$$

$$\Phi_{B_n}^y(x, y) = B_n(x, y + 1) - B_n(x, y). \tag{3.26}$$

Magnitude of the gradient of the n-th base component is defined as,

$$H_n(x, y) = \sqrt{\Phi_{B_n}^x(x, y)^2 + \Phi_{B_n}^y(x, y)^2}. \tag{3.27}$$

Let $p(x, y)$ be the function that finds which base part has the maximum gradient

magnitude at (x, y). Then, it can be defined as,

$$p(x,y) = \underset{1 \leq n \leq N}{\arg\max}\ H_n(x,y) \tag{3.28}$$

Then the gradients of the fused base part can be given as,

$$\Phi_B^x(x,y) = \Phi_{B_{p(x,y)}}^x(x,y), \tag{3.29}$$

$$\Phi_B^y(x,y) = \Phi_{B_{p(x,y)}}^y(x,y), \tag{3.30}$$

where $\Phi_{B_{p(x,y)}}^x(x,y)$ and $\Phi_{B_{p(x,y)}}^y(x,y)$ denote the x and y gradients of the base component with index $p(x,y)$ at pixel location (x,y). Finally, using the reconstruction method discussed above, the fused base part is reconstructed from the gradients.

In the experiments, N is equal to 2. However, this maximum absolute gradient fusion method is easily extended to any number of source images, as demonstrated in [96].

Maximum absolute gradient fusion rule is a simple yet effective method for the fusion of the base parts. Although most of the sharper details are in the detail components, some detail remain in base components. Many prior IR and visible image fusion methods based on MSD take the average of the base parts in the intensity domain, which leads to more smoothing in the fused base component. By utilizing the maximum absolute fusion strategy, we aim to better retain the remaining details in the base parts as well.

3.4.5 Two-scale image reconstruction

In the reconstruction step, the gradients of the fused base part and the detail part are added together to get the gradients of the fused image.

$$\Phi^x(x,y) = \Phi_B^x(x,y) + \Phi_D^x(x,y) \tag{3.31}$$

$$\Phi^y(x,y) = \Phi_B^y(x,y) + \Phi_D^y(x,y) \tag{3.32}$$

Finally, the image is reconstructed from the fused gradients $\Phi^x(x,y)$ and $\Phi^y(x,y)$ using the reconstruction algorithm elaborated in Section 3.2.

3.5 Summary

In this chapter, we presented the proposed gradient-domain fusion framework for IR/visible image pairs. The proposed method is aimed to combine the advantages of image decomposition and gradient domain-based fusion.

In the proposed method, the source images are first decomposed in to base and detail layers by utilizing a guided filter-based image decomposition scheme. In order to increase the retention of residual low-frequency information in base layers, we propose fusing base layers in the gradient domain by choosing the maximum absolute gradient. The gradients of the detail layers are fused using a weighted average scheme, where the weight maps are calculated using saliency maps. To obtain the respective saliency maps of the source images, we propose using the first convolutional layer of the VGG19 model. This enables utilizing the features learned by pre-trained CNN models with minimal increase in computational complexity. Prior to the fusion of the detail layers in the gradient domain, a guided filter-based image enhancement procedure is utilized in order to enhance the important information in the detail components.

The proposed method was evaluated against the state-of-the-art IR/visible fusion methods in terms of quantitative performance, qualitative performance and computational complexity. These experiments are presented in the next chapter.

Chapter 4

Experimental Results and Evaluation

4.1 Introduction

In this chapter, we analyze the effectiveness of the proposed method by comparing its qualitative performance, quantitative performance and computational complexity with that of the state-of-the-art methods in IR-visible image fusion. For this purpose, we use 45 IR-visible image pairs from the popular TNO image fusion dataset [109,110]. The TNO dataset was chosen for this evaluation since it is a well-known benchmarking dataset in IR-visible image fusion literature.

4.1.1 Other methods for comparison

The fusion performance of the proposed method is compared with 15 representative methods, namely, DLF [107], GFCE [111], Hybrid MSD [89], LatLRR [112], MD-LatLRR [113], MST-SR [51], Paul et al. [96], NSCT-SR [51], ResNet [114], RP-SR [51], DenseFuse [71], DIVFusion [69], NestFuse [72], PIAFusion [115] and SeAFusion [116]. From these 15 methods, the first ten are conventional methods whereas the remaining five are deep learning-based methods. These methods were chosen based on their fusion performance as well as the code availability. Furthermore, the five deep-learning based methods were chosen by selecting those that were published most recently, and trained on datasets other than the TNO dataset.

The methods MST-SR, NSCT-SR and RP-SR are all sparse representation based, and they were proposed as generic frameworks for any image fusion task [51]. DLF

[107] and ResNet [114] methods use pre-trained CNNs for generating the fusion weights with DLF using the VGG19 network whereas ResNet uses the ResNet50 network.

LatLRR method utilizes latent low-rank representation for two-level image decomposition and fuses the low-rank and salient parts separately by using hand-picked fusion rules [112]. MDLatLRR extends the same concept further by proposing a multi-level decomposition. Both GFCE and Hybrid MSD are based on hybrid multi-scale decompositions [113]. GFCE [111] uses two guided filters for its decomposition scheme whereas Hybrid MSD [89] uses Gaussian and bilateral filters.

Paul's method [96] is a gradient-based fusion scheme proposed for multi-focus and multi-exposure image fusion. As mentioned before, the method utilizes maximum gradient magnitude as the fusion rule.

Among the deep learning-based methods, DenseFuse [71], DIVFusion [69] and NestFuse [72] are encoder-decoder architectures. DenseFuse has a fairly simple architecture whereas DIVFusion has different modules for contrast enhancement and gradient retention. SeAFusion [116] is devised for real-time fusion and PIAFusion [115] emphasizes on extreme illumination conditions.

In the DenseFuse model, the deep features extracted by the encoder part are fused either via addition or by taking L_1 norm and averaging them. The reported subjective and objective fusion results are based on the averaging L_1 norm strategy. For the NestFuse model, [72] proposed three choices for *channel attention*; average pooling, max pooling, and nuclear pooling. The concept of *attention* in computer vision emulates the way human visual system draws our attention towards certain parts or objects of an image disregarding the other regions of the image. Attention mechanisms try to mimic this process by weighing the deep features in an adaptive manner along the spatial dimensions and channel dimension. For details on attention in computer vision, interested reader is referred to [117]. In this work, the demonstrated NestFuse results were obtained using the nuclear pooling channel attention mechanism.

4.1.2 Objective Fusion Metrics

In the image fusion literature, a vast variety of image quality metrics have been used to measure the objective fusion performance. These metrics can be categorized in to 4 groups, namely, (1) information theory-based metrics, (2) image feature-based metrics, (3) structural similarity-based metrics, and (4) human perception inspired

metrics [95].

Liu et al. performed a comparative study of 12 such metrics over 6 multiresolution image fusion algorithms on IR-visible image fusion data [95]. However, currently there is no consensus about a single metric that is best suited to measure the IR and visible image fusion performance quantitatively. Therefore, 10 commonly used metrics representing all quantitative metric categories were chosen in this study, as well as a qualitative analysis (section 4.3) based on the visual quality of the fused images. The chosen metrics are, average gradient (AG), edge intensity (EI), normalized mutual information (NMI), Cvejie's metric (Q_C), Chen-Varshney metric (Q_{CV}), standard deviation (SD), spatial frequency (SF), structural similarity index measure (SSIM), Tsallis entropy (TE), and visual information fidelity (VIF).

Here, NMI and TE are information theory-based metrics. AG, EI, SF, SD are feature-based metrics. Q_{CV} and VIF are human perception inspired metrics, whereas Q_C and SSIM are structural similarity-based metrics. For all the metrics except Q_{CV}, a larger value indicates better performance. Further details about each metric is available in Appendix A.

4.2 Quantitative Performance Comparison

Table 4.1 displays the average metric values obtained by each method on the TNO dataset. As seen in Table 4.1, the proposed method demonstrates the best quantitative performance with respect to VIF metric, and ranks a close second in terms of the metrics AG, EI, SD and SF.

MDLatLRR achieves the best fusion performance with respect to 3 metrics namely, AG, EI and SF. These 3 metrics reflect the sharpness and texture richness of an image. This top performance is also reflected in qualitative results as seen in Figure 4.1.

PIAFusion attained the best performance in terms of the information theory-based metrics Q_C and TE. However, it can be noted that the background details or important targets are less clear in some fused images of Figure 4.1.

DLF, NSCT-SR, DenseFuse and DIVFusion demonstrate the best objective fusion performance with respect to one fusion metric each. Those metrics are SSIM, NMI, Q_C and SD respectively. Among these four methods, DLF outputs lack contrast and NSCT-SR outputs have visible artifacts as seen in Figure 4.1. While the DIVFusion fused images demonstrate higher contrast compared to many other methods, some lack clarity near the interesting targets.

Table 4.1: Comparison with 15 competing fusion methods. All metrics are averages over 45 image pairs from the TNO dataset [110]. Red, blue and green denote the top 3 values in order.

Method	AG	EI	NMI	Q_C	Q_{CV}	SD	SF	SSIM	TE	VIF
DLF [107]	2.3818	23.4781	0.3847	0.6689	421.54	26.4360	6.4787	1.5579	2178.12	0.3335
GFCE [111]	5.7949	55.8209	0.2665	0.5864	567.64	43.3454	15.3098	1.2039	1191.28	0.7970
Hybrid MSD [89]	4.2266	40.9199	0.2690	0.6738	470.17	38.7859	11.5583	1.2818	3372.74	0.5588
LatLRR [112]	6.8671	67.4776	0.3616	0.6771	555.52	43.9458	8.3078	1.4442	596.31	0.8419
MDLatLRR [113]	7.9492	78.2892	0.3903	0.6989	592.88	41.9134	20.3672	1.4292	489.75	1.1881
MST-SR [51]	4.0857	40.0146	0.2192	0.5924	590.89	44.6009	11.0075	1.1456	369.19	0.5499
NSCT-SR [51]	4.6612	46.8595	0.5750	0.6976	776.41	41.2324	11.7778	1.2987	397.91	0.3668
ResNet [114]	2.3274	22.9891	0.3512	0.6583	428.02	26.3615	6.1582	1.5579	2175.36	0.3293
RP-SR [51]	4.6974	44.0496	0.3224	0.6509	699.82	44.4576	14.0682	1.3260	369.79	0.4627
DenseFuse [71]	3.4564	35.2579	0.4225	0.5762	264.97	40.7409	9.3737	1.3906	4186.23	0.4746
DIVFusion [69]	5.1124	49.7579	0.3381	0.6408	704.15	54.3429	12.7983	1.1942	28801.05	0.6284
NestFuse [72]	3.6815	36.4687	0.4791	0.6760	276.91	42.3856	9.8761	1.4407	1659.59	0.4988
PIAFusion [115]	4.0011	39.0417	0.4925	0.7827	283.01	41.5968	10.5157	1.4248	58786.91	0.4401
SeAFusion [116]	4.4874	45.5857	0.4041	0.6064	266.05	44.0360	11.2058	1.3584	11484.47	0.5777
Paul et al. [96]	5.4514	52.6801	0.2974	0.5984	495.26	40.8194	13.6648	1.1813	40326.84	0.9073
Proposed	7.2751	70.5310	0.2324	0.5706	505.67	44.6256	17.6140	1.1320	5108.44	1.3609

According to Table 4.1, the top performances of the 10 metrics are spread across various methods. However, it should be noted that the proposed method ranks in the top 3 for 5 metrics. The metrics for which this was observed includes all the feature based metrics and the human perception-based metric VIF. This indicates that the proposed method has superior performance in terms of sharpness, contrast and fidelity of the images.

4.3 Qualitative Performance Comparison

In this section, the proposed method is compared against other methods using a qualitative comparison of the visual quality of the fused images with respect to information retention, contrast and image brightness.

Figures 4.1 and 4.2 juxtapose the fusion results of the selected methods against those of the proposed method for 7 image pairs from the TNO dataset. Further, in Figures 4.3 and 4.4 some cropped areas from the source images and the fused images are used to visualize the performance of the proposed method with respect to information and contrast retention. For better viewing, source images and the output

Figure 4.1: Fused results of image sequences (a) '*Kaptein 1123*', (b) '*Kaptein 1654*', (c) '*soldier in trench 2*', (d) '*man in doorway*', (e) '*Marne 04*', (f) '*Marne 24*', and (g) '*soldier behind smoke 1*', and for DLF, GFCE, HYbrid MSD, LatLRR, MDLatLRR, MST-SR, NSCT-SR and ResNet. The top row and the second row depicts the source IR image and the visible image for each image pair.

Figure 4.2: Fused results of image sequences (a) '*Kaptein 1123*', (b) '*Kaptein 1654*', (c) '*soldier in trench 2*', (d) '*man in doorway*', (e) '*Marne 04*', (f) '*Marne 24*', and (g) '*soldier behind smoke 1*', and for RP-SR, DenseFuse, DIVFusion, NestFuse (nuclear), PIAFusion, SeAFusion, Paul's method and the proposed method. The top row and the second row depicts the source IR image and the visible image for each image pair.

images from each comparison method have been shared in a public repository.[1]

In (a) to (c) source image pairs of Figure 4.1, the IR images contain most of the target information and almost completely lack the background textural details which are found in the visible images. In sequences (d) to (g), the IR images also contain important background contextual information in addition to the target details.

For these 7 examples of Figure 4.1, the DLF method fairly retains most information but the fused images are overall darker and lack contrast which lead to some targets not being sufficiently clear. NSCT-SR outputs display visible artifacts whereas RP-SR fused images have over-exposed areas.

GFCE method has improved illumination in the fused outputs compared to the source images. However, it has also enhanced the noise in sequence (a) and has lost some textural information compared to the other methods. Outputs of LatLRR method are overexposed and it has missed the target information of sequence (g) from the fused image. Compared to this, MDLatLRR which is the improved version of LatLRR, performs the quite well in terms of information retention. MST-SR lacks sharpness in the fused images whereas ResNet outputs are lacking in contrast.

Regarding deep learning-based methods, DenseFuse outputs highlight the targets from IR images well, but lack some background information from the visible images. Fused images of the DIVFusion model have overexposed areas which cause the target to be less contrasted from the background. PIAFusion performs well in many sequences, but the fused output for sequence (g) completely lacks the target information. Similarly, SeAFusion model also fails to highlight the target in example (g). NestFuse performs relatively better compared to DIVFusion, PIAFusion, and SeAFusion models. However, it has difficulty in highlighting the target of sequence (g) and retaining background details in (e). Theoretically, the deep learning models failing in sequence (g) can be explained by the representativeness of the data used in training. It is highly likely that the data used for training these models did not contain sufficient samples where targets were obstructed by smoke.

In terms of detail retention, the MDLatLRR method and Paul's method perform quite well on all the example sequences. Further, it can be noted that MDLatLRR outputs have a halo effect around important targets, whereas the outputs of Paul's method lack in contrast and background textural details. The proposed fusion method succeeds to retain the IR target information and background details while attaining improved contrast and brightness.

[1] `https://github.com/sndnshr/IR-Visible-Image-Fusion-in-the-Gradient-Domain`

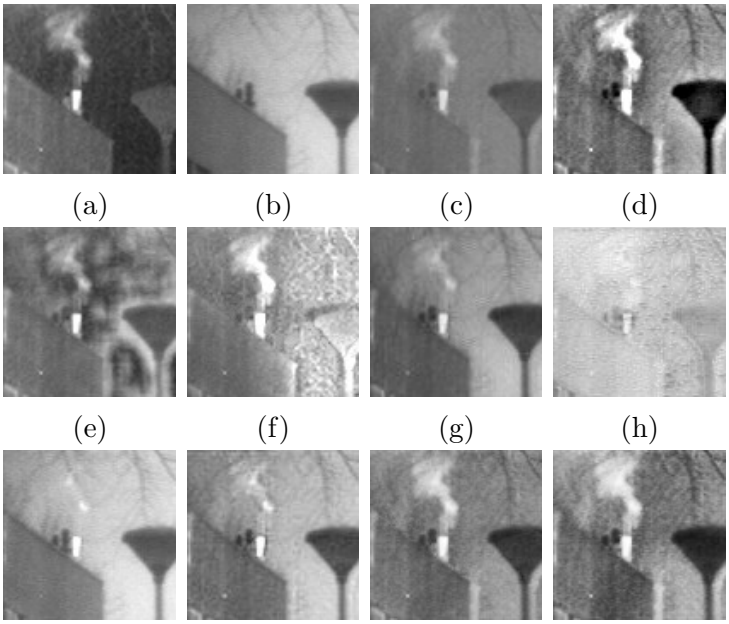


Figure 4.3: Crops from fusion results of '*Kaptein 1123*' image sequence corresponding to the red box on the source images in Figure 4.1 (a) IR image (b) Visible image (c) LatLRR (d) MDLatLRR (e) MST-SR (f) RP-SR (g) DenseFuse (h) DIVFusion (i) PIAFusion (j) SeAFusion (k) Paul's method (l) Proposed.

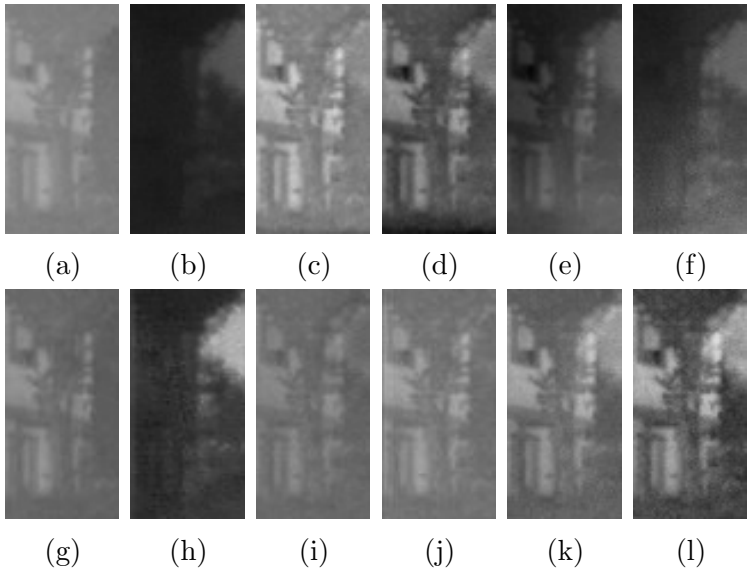


Figure 4.4: Crops from fusion results of '*Marne 04*' image sequence corresponding to the red box on the source images in Figure 4.1 (a) IR image (b) Visible image (c) LatLRR (d) MDLatLRR (e) MST-SR (f) RP-SR (g) DenseFuse (h) DIVFusion (i) PIAFusion (j) SeAFusion (k) Paul's method (l) Proposed.

Table 4.2: Average runtime comparison on the TNO image fusion dataset.

Method	Avg runtime (s) without GPU	Avg runtime (s) with GPU
DLF [107]	17.1342	
GFCE [111]	2.0125	
Hybrid MSD [89]	7.8588	
LatLRR [112]	234.4919	
MDLatLRR [113]	400.4850	
MST-SR [51]	0.98576	
NSCT-SR [51]	111.2974	
ResNet [114]	7.3620	
RP-SR [51]	0.9854	
DenseFuse - addition [71]	2.4210	1.2910
DenseFuse - average L_1 norm [71]	5.7147	5.8957
DIVFusion [69]	60.5547	2.4392
NestFuse - avg [72]	22.5683	9.1619
NestFuse - max [72]	22.6922	8.8633
NestFuse - nuclear [72]	53.1898	37.4816
PIAFusion [115]	10.5884	0.8126
SeAFusion [116]	7.9362	0.1363
Paul et al. [96]	0.8430	
Proposed	1.1333	

Figure 4.3 depicts the cropped portions corresponding to the red box on sequence (a) IR image of Figure 4.1. In this example, MDLatLRR, Paul's method, and the proposed method perform better in retaining the most information from the source regions.

The lack of contrast and exposure issues of the fused images can be further observed in Figure 4.4. In this example, the outputs of RP-SR and DIVFusion demonstrate underexposed fused images and almost completely lack the information in the IR image. Compared to other methods, MDLatLRR, Paul's method and the proposed method perform well in this example. It can be seen that the proposed method demonstrates the best fusion result with good exposure and better detail retention.

In conclusion, compared to the competing methods, the proposed method optimally preserves the target and background information from the source images while also improving the contrast and the overall brightness of the fused images.

4.4 Computational Complexity Comparison

Table 4.2 displays the average runtime in seconds on the TNO dataset [110]. The conventional methods were evaluated using their MATLAB implementations. The deep learning-based methods were evaluated using their Pytorch or Tensorflow implementations.

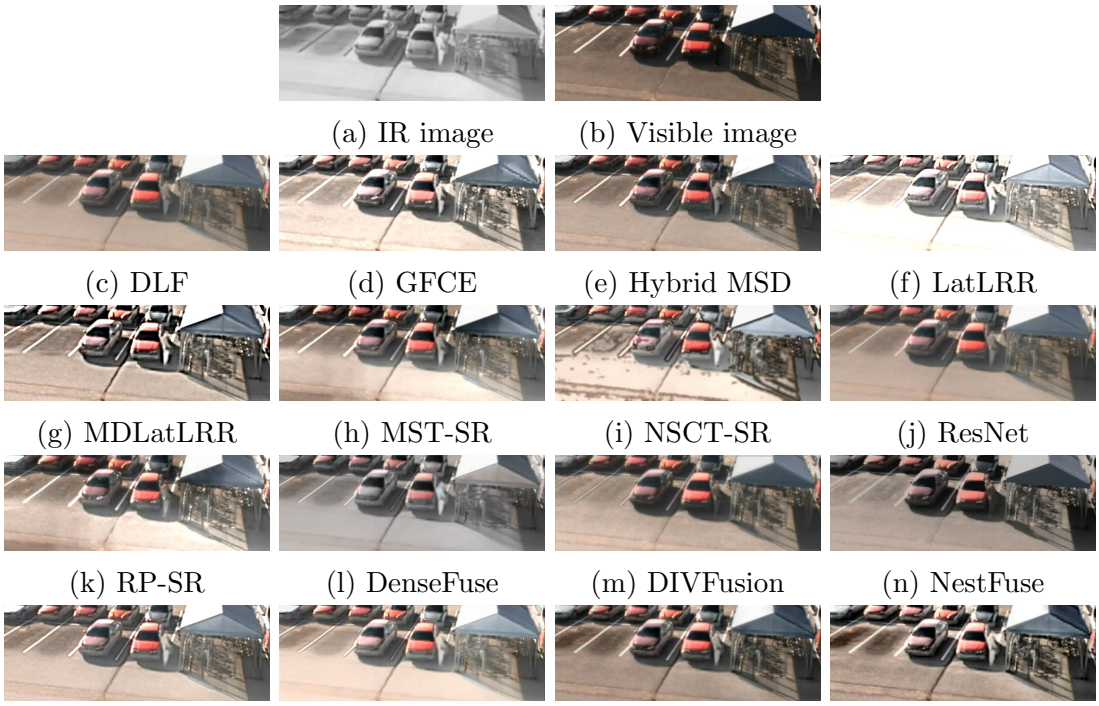
The second column of Table 4.2 shows the runtime performance evaluated on an Intel i5 PC with 16 GB RAM and Windows 10 operating system without GPUs. Here, the deep learning-based methods were tested using a Python 3.6 Anaconda environment.

The third column lists the run times of the deep learning-based methods on a computing platform with GPU support. Here, the experiments were performed on a Broadwell node of the Cedar cluster located at Simon Fraser University, Canada [118]. Each method was tested on a Broadwell node which consisted of an NVIDIA Tesla P100 PCIe 12 GB GPU, 4 CPU cores and 16 GB RAM.

As evident from this juxtaposition, the proposed method is among the fastest methods. While RP-SR [51] have lower computational time than the proposed method, the fused images are overexposed in some regions and lack clarity around targets. Even though MST-SR [51] takes lesser computational time, the proposed method outperforms it in the quantitative analysis. Paul et al. [96] method is also $\approx 25\%$ faster, but lacks the detail and contrast produced by the proposed fusion scheme.

Among the 5 deep learning-based methods, DenseFuse is the fastest when evaluated without GPUs. Regarding NestFuse, the three channel attention mechanisms are indistinguishable with respect to quantitative and qualitative fusion performance. However, in terms of computational efficiency, NestFuse with the nuclear pooling channel attention mechanism ranks lower compared to the other attention schemes. The SeAFusion model which was proposed as a real-time fusion network achieves a significantly fast computation time with GPU support.

From the above analysis it can be concluded that the proposed method performs faster than most existing methods without the need for additional hardware such as GPUs, while also achieving significant quantitative and qualitative fusion performance.



(a) IR image (b) Visible image

(c) DLF (d) GFCE (e) Hybrid MSD (f) LatLRR

(g) MDLatLRR (h) MST-SR (i) NSCT-SR (j) ResNet

(k) RP-SR (l) DenseFuse (m) DIVFusion (n) NestFuse

(o) PIAFusion (p) SeAFusion (q) Paul et al. (r) Proposed

Figure 4.5: Fusion results for '*carShadow*' image pair from the VIFB dataset.

4.5 Fusion of color images

As mentioned in Section 3.4, when visible images are color, the common approach is to fuse the luminance channel of the color image with the grayscale IR image. For testing this approach, we chose the sequences with RGB visible images from the VIFB dataset [119].

Figure 4.5 depicts the fusion outputs for the *carShadow* sequence in the VIFB dataset. With outputs from DLF, ResNet, DIVFusion and NestFuse the target (man near the red car) is not highlighted sufficiently. Output of NSCT-SR demonstrates artifacts whereas LatLRR output is overexposed. Among the considered methods, MDLatLRR, DenseFuse and the proposed method perform better in highlighting the human target compared to the others.

It is evident that when the visible image is color, the fused images demonstrate subjective fusion performance which is consistent with the same method's performance on grayscale image pairs. This phenomenon is expected since important target and background details are contained in the luminance channel, whereas the chorminance channels contain color information. This is also the reason why the objective

image fusion metrics only operate on the luminance channel.

4.6 Summary

Qualitative and quantitative comparison with the state-of-the-art methods demonstrate that the proposed method outperforms most comparative methods visually and objectively, while also achieving significantly faster computational time without requiring additional hardware such as GPUs.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Image fusion is the process by which multiple images of a scene are combined to produce a single more informative image. Among the many categories of image fusion, IR-visible image fusion is highly desirable due to its capability in improving the visual understanding of low-visibility scenes.

In this thesis, we proposed a novel IR-visible image fusion method based on the gradient domain. In order to achieve the significant subjective and objective fusion performance in conjunction with the low computational complexity, we employed a two-scale strategy for the fusion in the gradient domain. The base parts are fused in the gradient domain by choosing the maximum absolute gradient of the two source images. In contrast to this, the detail parts are fused using a saliency-based strategy. For detecting the salient regions, we employed a simple approach using the pre-trained CNN VGG-19 which was trained for the ImageNet classification challenge. In this approach, we utilize the activations of the very first convolutional layer and obtain the L_1 norm across the channel dimension to generate a saliency map. The two saliency maps will then be normalized so that they sum to 1. These normalized saliency maps will be the weight maps for fusing the gradients of the detail parts. Also, prior to fusion, we utilized guided filter-based image enhancement on the detail parts. Finally, the fused image is reconstructed from the sum of base part and detail part gradients.

Utilizing the TNO dataset, we demonstrated that the proposed method shows comparative fusion performance to state-of-the-art methods in terms of objective fusion criteria as well as subjective evaluation. Since there is no consensus regarding

the best-suited objective metric for IR-visible image fusion, many different metrics are utilized in the literature. Hence, we chose 10 representative and commonly used metrics for our performance evaluation. From the quantitative results we showed that the proposed method demonstrates significantly high performance in all the feature-based metrics as well as the top performance in the human perception-inspired metric VIF.

In addition to this, we evaluated the computational complexity of the proposed method against the state-of-the-art methods in literature. It was evident that the proposed approach has significantly lesser run-time which can be achieved without additional hardware support.

5.2 Future Work

Fusion in the gradient domain can be applied to MMIF and specifically for fusion of three-dimensional (3D) medical images. Many medical imaging modalities such as CT, MRI are 3D images. The currently common approach for fusion of these volumetric images is to fuse them per 2D slice. The gradient domain method can be explored as a 3D fusion approach by extending the decomposition and reconstruction algorithms to 3D. Also, it is necessary to find appropriate fusion rules as well.

Since deep learning-based methods have shown immense proficiency in many computer vision tasks, an interesting future avenue would be to explore how to design more efficient lightweight models for fusion while retaining high fusion performance. Computational complexity is a crucial problem since tracking and surveillance applications require processing of videos.

Furthermore, application-oriented deep learning-based fusion would be an interesting problem to work on. Here, the idea is that the model will learn features that are optimal for the downstream applications such as detection or surveillance.

Appendix A

Objective Fusion Metrics

1. The average gradient (AG) metric quantifies the amount of gradient information in an image, which represents the sharpness of details and richness of texture. The definition of AG is given by,

$$AG = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\frac{(I(i,j) - I(i+1,j))^2 + (I(i,j) - I(i,j+1))^2}{2}} \quad \text{(A.1)}$$

   where $I(i,j)$ is the grayscale value of pixel (i,j), for an image of size M by N. The larger the AG value, the more gradient information is contained in the fused image.

2. Edge intensity (EI) of the fused image F can be measured as follows,

$$EI = \frac{1}{M \times N} \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} S_x(i,j)^2 + S_y(i,j)^2} \quad \text{(A.2)}$$

   where S_x and S_y are the results from convolving with the Sobel operator.

   A larger EI value signifies higher image quality in terms of clarity and image sharpness.

3. Normalized mutual information (NMI) is formulated using mutual information, which is measure of the mutual dependence of two random variables. The

Hossny et al. [120] definition of NMI can be given as,

$$NMI = 2\left[\frac{MI(A,F)}{H(A)+H(F)} + \frac{MI(B,F)}{H(B)+H(F)}\right] \tag{A.3}$$

where, $MI(A,F)$ and $MI(B,F)$ are the mutual information between source images A, B with fused image F. Also, $H(A), H(B)$ and $H(F)$ denote the marginal entropy of A, B and F respectively.

4. Cvejie's metric (Q_C) can be defined as [121],

$$Q_c = \sum_{w \in W} sim(A,B,F|w)Q(A|F,w) + (1 - sim(A,B,F|w))Q(B,F|w) \tag{A.4}$$

$$sim(A,B,F|w) = \begin{cases} 0, & \text{if } \sigma(A,B,F) < 0, \\ \sigma, & \text{if } 0 \leq \sigma(A,B,F) \leq 1, \\ 1, & \text{if } \sigma(A,B,F) > 1. \end{cases} \tag{A.5}$$

$$\sigma(A,B,F) = \frac{\sigma_{AF}}{\sigma_{AF} + \sigma_{BF}} \tag{A.6}$$

5. Chen-Varshney metric (Q_{CV}) consists of five steps: extracting edge information, partitioning images in to local regions, calculating local region saliency, taking a similarity measure in the local region, and weighted summation of similarity measure over non-overlapping regions to obtain the final global quality measure [122].

6. Spatial frequency (SF) metric is based on image gradients and reflects the image detail and sharpness of texture. Its definition can be given as follows;

$$\begin{aligned} SF &= \sqrt{RF^2 + CF^2} \\ RF &= \sqrt{\sum_{i=1}^{M}\sum_{j=2}^{N}[F(i,j) - F(i,j-1)]^2} \\ CF &= \sqrt{\sum_{i=2}^{M}\sum_{j=1}^{N}[F(i,j) - F(i-1,j)]^2} \end{aligned} \tag{A.7}$$

where RF is the row frequency and CF is the column frequency of the fused

image F. An image with high SF has sharp edges and rich texture information, which the human visual perception is sensitive to.

7. Standard deviation (SD) metric is a measure of the contrast of the fused image. It is mathematically defined as follows.

$$SD = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j)-\mu)^2}, \tag{A.8}$$

where F is the fused image and μ is its mean value. Due to the sensitivity of the human visual system to contrast, regions with high contrast tend to attract attention. A fused image with high contrast often results in high SD.

8. Structural similarity index measure (SSIM) models the distortion of an image using three aspects: luminance, contrast, and correlation. The three components are combined as follows [95].

$$SSIM(X,F) = \left(\frac{2\mu_X\mu_F + c_1}{\mu_X^2 + \mu_F^2 + c_1}\right)^{\alpha}\left(\frac{2\sigma_X\sigma_F + c_2}{\sigma_X^2 + \sigma_F^2 + c_2}\right)^{\beta}\left(\frac{\sigma_{XF} + c_3}{\sigma_X\sigma_F + c_3}\right)^{\gamma} \tag{A.9}$$

Here, μ_X, μ_F are the averages, σ_X, σ_F and σ_{XF} are the variances and covariance of the source image X and fused image F. The parameters α, β and γ are used to adjust the relative importance of the three components. c_1, c_2 and c_3 are small constants used to ensure numerical stability.

Larger values of SSIM indicate that the two images considered are more similar in terms of brightness, contrast and structure. SSIM for fusion will be given by,

$$SSIM = SSIM(A,F) + SSIM(B,F) \tag{A.10}$$

where A and B denote the source images and F denotes the fused image.

9. Tsallis entropy (TE) is defined as [123],

$$Q_{TE}^{q} = I^q(A,F) + I^q(B,F). \tag{A.11}$$

Here, $I^q(A,F)$ and $I^q(B,F)$ is found via

$$I^q(X,F) = \frac{1}{1-q}\left(1 - \sum_{i,j} \frac{h_{XF}(i,j)^q}{h_F(j)h_X(i)^{q-1}}\right). \tag{A.12}$$

where $h_{XF}(i,j)$, $h_X(i)$ and $h_F(j)$ denote the joint and marginal histograms of respective images, q is a real value and $q \neq 1$.

10. Visual information fidelity (VIF) metric measures the information fidelity by computing the distortion between the source images and the fusion result. It can be calculated using the procedure detailed in [124].

Bibliography

[1] Y. Li, M. Liu, and K. Han, "Overview of multi-exposure image fusion," in *2021 International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*, pp. 196–198, 2021, doi:10.1109/ICEIB53692.2021.9686453.

[2] Y. Liu, L. Wang, J. Cheng, C. Li, and X. Chen, "Multi-focus image fusion: A survey of the state of the art," *Information Fusion*, vol. 64, pp. 71–91, 2020, doi:10.1016/j.inffus.2020.06.013.

[3] M. A. Azam, K. B. Khan, S. Salahuddin, E. Rehman, S. A. Khan, M. A. Khan, S. Kadry, and A. H. Gandomi, "A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics," *Computers in Biology and Medicine*, vol. 144, p. 105253, 2022, doi:10.1016/j.compbiomed.2022.105253.

[4] H. Ghassemian, "A review of remote sensing image fusion methods," *Information Fusion*, vol. 32, pp. 75–89, 2016, doi:10.1016/j.inffus.2016.03.003.

[5] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Information Fusion*, vol. 62, pp. 110–120, 2020, doi:10.1016/j.inffus.2020.04.006.

[6] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Machine vision and applications*, vol. 25, pp. 245–262, 2014, doi:10.1007/s00138-013-0570-5.

[7] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Information Fusion*, vol. 45, pp. 153–178, 2019, doi:10.1016/j.inffus.2018.02.004.

[8] F. Omri, S. Foufou, and M. Abidi, "NIR and visible image fusion for improving face recognition at long distance," in *Image and Signal Processing*, pp. 549–557, Springer International Publishing, 2014, doi:10.1007/978-3-319-07998-1_63.

[9] H. Hariharan, A. Koschan, B. Abidi, A. Gribok, and M. Abidi, "Fusion of visible and infrared images using empirical mode decomposition to improve face recognition," in *2006 International Conference on Image Processing*, pp. 2049–2052, 2006, doi:10.1109/ICIP.2006.312860.

[10] J. Heo, S. Kong, B. Abidi, and M. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pp. 122–122, 2004, doi:10.1109/CVPR.2004.351.

[11] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image and Vision Computing*, vol. 24, no. 7, pp. 727–742, 2006, doi:10.1016/j.imavis.2006.01.017.

[12] W. He, W. Feng, Y. Peng, Q. Chen, G. Gu, and Z. Miao, "Multi-level image fusion and enhancement for target detection," *Optik*, vol. 126, no. 11, pp. 1203–1208, 2015, doi:10.1016/j.ijleo.2015.02.092.

[13] J. C. Castillo, A. Fernández-Caballero, J. Serrano-Cuerda, M. T. López, and A. Martínez-Rodrigo, "Smart environment architecture for robust people detection by infrared and visible video fusion," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, pp. 223–237, 2017, doi:10.1007/s12652-016-0429-5.

[14] E. Fendri, R. R. Boukhriss, and M. Hammami, "Fusion of thermal infrared and visible spectra for robust moving object detection," *Pattern Analysis and Applications*, vol. 20, pp. 907–926, 2017, doi:10.1007/s10044-017-0621-z.

[15] S. R. Schnelle and A. L. Chan, "Enhanced target tracking through infrared-visible image fusion," in *14th International Conference on Information Fusion*, pp. 1–8, IEEE, 2011.

[16] K. S. Kumar, G. Kavitha, R. Subramanian, and G. Ramesh, "Visual and thermal image fusion for uav based target tracking," in *MATLAB–A Ubiquitous Tool for the Practical Engineer*, ch. 16, IntechOpen, 2011, doi:10.5772/23996.

[17] G. Bhatnagar and Z. Liu, "A novel image fusion framework for night-vision navigation and surveillance," *Signal, Image and Video Processing*, vol. 9, pp. 165–175, 2015, doi:10.1007/s11760-014-0740-6.

[18] N. Paramanandham and K. Rajendiran, "Multi sensor image fusion for surveillance applications using hybrid image fusion algorithm," *Multimedia Tools and Applications*, vol. 77, pp. 12405–12436, 2018, doi:10.1007/s11042-017-4895-3.

[19] J. Tao, Y. Cao, M. Ding, and Z. Zhang, "Visible and infrared image fusion-based image quality enhancement with applications to space debris on-orbit surveillance," *International Journal of Aerospace Engineering*, vol. 2022, 2022, doi:10.1155/2022/6300437.

[20] X. Chang, L. Jiao, F. Liu, and F. Xin, "Multicontourlet-based adaptive fusion of infrared and visible remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 3, pp. 549–553, 2010, doi:10.1109/LGRS.2010.2041323.

[21] A. Ahmad, M. M. Riaz, A. Ghafoor, T. Zaidi, *et al.*, "An improved infrared/visible fusion for astronomical images," *Advances in Astronomy*, vol. 2015, 2015, doi:10.1155/2015/203872.

[22] L. Han, B. Wulie, Y. Yang, and H. Wang, "Direct fusion of geostationary meteorological satellite visible and infrared images based on thermal physical properties," *Sensors*, vol. 15, no. 1, pp. 703–714, 2015, doi:10.3390/s150100703.

[23] M. Eslami and A. Mohammadzadeh, "Developing a spectral-based strategy for urban object detection from airborne hyperspectral tir and visible data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1808–1816, 2016, doi:10.1109/JSTARS.2015.2489838.

[24] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983, doi:10.1109/TCOM.1983.1095851.

[25] D. Bulanon, T. Burks, and V. Alchanatis, "Image fusion of visible and thermal images for fruit detection," *Biosystems Engineering*, vol. 103, no. 1, pp. 12–22, 2009, doi:10.1016/j.biosystemseng.2009.02.009.

[26] A. V. Vanmali and V. M. Gadre, "Visible and NIR image fusion using weight-map-guided laplacian–gaussian pyramid for improving scene visibility," *Sādhanā*, vol. 42, pp. 1063–1082, 06 2017, doi:10.1007/s12046-017-0673-1.

[27] Z. Liu, K. Tsukada, K. Hanasaki, Y.-K. Ho, and Y. Dai, "Image fusion by using steerable pyramid," *Pattern Recognition Letters*, vol. 22, no. 9, pp. 929–939, 2001, doi:10.1016/S0167-8655(01)00047-2.

[28] G. Liu, Z. Jing, S. Sun, J. Li, Z. Li, and H. Leung, "Image fusion based on expectation maximization algorithm and steerable pyramid," *Chinese Optics Letters*, vol. 2, no. 7, pp. 386–389, 2004.

[29] H. Jin, L. Jiao, F. Liu, and Y. Qi, "Fusion of infrared and visual images based on contrast pyramid directional filter banks using clonal selection optimizing," *Optical Engineering*, vol. 47, no. 2, p. 027002, 2008, doi:10.1117/1.2857417.

[30] H. Jin and Y. Wang, "A fusion method for visible and infrared images based on contrast pyramid with teaching learning based optimization," *Infrared Physics & Technology*, vol. 64, pp. 134–142, 2014, doi:10.1016/j.infrared.2014.02.013.

[31] L. Chipman, T. Orr, and L. Graham, "Wavelets and image fusion," in *Proceedings., International Conference on Image Processing*, vol. 3, pp. 248–251 vol.3, 1995, doi:10.1109/ICIP.1995.537627.

[32] Y. Niu, S. Xu, L. Wu, and W. Hu, "Airborne infrared and visible image fusion for target perception based on target region segmentation and discrete wavelet transform," *Mathematical Problems in Engineering*, vol. 2012, 2012, doi:10.1155/2012/275138.

[33] L. Zhan, Y. Zhuang, and L. Huang, "Infrared and visible images fusion method based on discrete wavelet transform," *Journal of Computers*, vol. 28, no. 2, pp. 57–71, 2017, doi:10.3966/199115592017042802005.

[34] J. Saeedi and K. Faez, "Infrared and visible image fusion using fuzzy logic and population-based optimization," *Applied Soft Computing*, vol. 12, no. 3, pp. 1041–1054, 2012, doi:10.1016/j.asoc.2011.11.020.

[35] Y. Zuo, J. Liu, G. Bai, X. Wang, and M. Sun, "Airborne infrared and visible image fusion combined with region segmentation," *Sensors*, vol. 17, no. 5, p. 1127, 2017, doi:10.3390/s17051127.

[36] F. Meng, M. Song, B. Guo, R. Shi, and D. Shan, "Image fusion based on object region detection and non-subsampled contourlet transform," *Computers & Electrical Engineering*, vol. 62, pp. 375–383, 2017, doi:10.1016/j.compeleceng.2016.09.019.

[37] S. Yin, L. Cao, Q. Tan, and G. Jin, "Infrared and visible image fusion based on NSCT and fuzzy logic," in *2010 IEEE International Conference on Mechatronics and Automation*, pp. 671–675, IEEE, 2010, doi:10.1109/ICMA.2010.5588318.

[38] J. Cai, Q. Cheng, M. Peng, and Y. Song, "Fusion of infrared and visible images based on nonsubsampled contourlet transform and sparse K-SVD dictionary learning," *Infrared Physics & Technology*, vol. 82, pp. 85–95, 2017, doi:10.1016/j.infrared.2017.01.026.

[39] H. Li, H. Qiu, Z. Yu, and Y. Zhang, "Infrared and visible image fusion scheme based on NSCT and low-level visual features," *Infrared Physics & Technology*, vol. 76, pp. 174–184, 2016.

[40] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM transactions on graphics (TOG)*, vol. 27, no. 3, pp. 1–10, 2008.

[41] D. P. Bavirisetti and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection," *Infrared Physics & Technology*, vol. 76, pp. 52–64, 2016.

[42] J. Hu and S. Li, "The multiscale directional bilateral filter and its application to multisensor image fusion," *Information Fusion*, vol. 13, no. 3, pp. 196–206, 2012.

[43] B. K. S. Kumar, "Image fusion based on pixel significance using cross bilateral filter," *Signal, Image and Video Processing*, vol. 9, pp. 1193–1204, 07 2015, doi:10.1007/s11760-013-0556-9.

[44] J. Zhao, Q. Zhou, Y. Chen, H. Feng, Z. Xu, and Q. Li, "Fusion of visible and infrared images using saliency analysis and detail preserving based image decomposition," *Infrared physics & technology*, vol. 56, pp. 93–99, 2013, doi:10.1016/j.infrared.2012.11.003.

[45] Y. Jiang and M. Wang, "Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter," *IET Image Processing*, vol. 8, no. 3, pp. 183–190, 2014, doi:10.1049/iet-ipr.2013.0429.

[46] A. Toet and M. A. Hogervorst, "Multiscale image fusion through guided filtering," in *Target and Background Signatures II*, vol. 9997, p. 99970J, International Society for Optics and Photonics, SPIE, 2016, doi:10.1117/12.2239945.

[47] H.-M. Hu, J. Wu, B. Li, Q. Guo, and J. Zheng, "An adaptive fusion algorithm for visible and infrared videos based on entropy and the cumulative distribution of gray levels," *IEEE Transactions on Multimedia*, vol. 19, no. 12, pp. 2706–2719, 2017, doi:10.1109/TMM.2017.2711422.

[48] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, 2016, doi:10.1109/LSP.2016.2618776.

[49] X. Lu, B. Zhang, Y. Zhao, H. Liu, and H. Pei, "The infrared and visible image fusion algorithm based on target separation and sparse representation," *Infrared Physics & Technology*, vol. 67, pp. 397–407, 2014, doi:10.1016/j.infrared.2014.09.007.

[50] Y. Yao, P. Guo, X. Xin, and Z. Jiang, "Image fusion by hierarchical joint sparse representation," *Cognitive Computation*, vol. 6, no. 3, pp. 281–292, 2014, doi:10.1007/s12559-013-9235-y.

[51] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, pp. 147–164, 2015, doi:10.1016/j.inffus.2014.09.004.

[52] H. Mitchell, *Image Fusion Theories, Techniques and Applications*, Berlin, Heidelberg: Springer, 2010, doi:10.1007/978-3-642-11216-4.

[53] A. M. Yi Zhou and M. A. Omar, "Principal component analysis-based image fusion routine with application to automotive stamping split detection," *Research in Nondestructive Evaluation*, vol. 22, no. 2, pp. 76–91, 2011, doi:10.1080/09349847.2011.553348.

[54] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *2017 20th International Conference on Information Fusion (Fusion)*, pp. 1–9, 2017, doi:10.23919/ICIF.2017.8009719.

[55] H. Li, L. Liu, W. Huang, and C. Yue, "An improved fusion algorithm for infrared and visible images based on multi-scale transform," *Infrared Physics & Technology*, vol. 74, pp. 28–37, 2016, doi:10.1016/j.infrared.2015.11.002.

[56] S. S. Kumar and S. Muttan, "PCA-based image fusion," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII* (S. S. Shen and P. E. Lewis, eds.), vol. 6233, p. 62331T, International Society for Optics and Photonics, SPIE, 2006, doi:10.1117/12.662373.

[57] U. Patil and U. Mudengudi, "Image fusion using hierarchical pca.," in *2011 International Conference on Image Information Processing*, pp. 1–6, 2011, doi:10.1109/ICIIP.2011.6108966.

[58] Y. Zheng, E. A. Essock, and B. C. Hansen, "An advanced image fusion algorithm based on wavelet transform: incorporation with PCA and morphological processing," in *Image Processing: Algorithms and Systems III*, vol. 5298, pp. 177 – 187, International Society for Optics and Photonics, SPIE, 2004, doi:10.1117/12.523966.

[59] V. Naidu, "Hybrid DDCT-PCA based multi sensor image fusion," *Journal of Optics*, vol. 43, pp. 48–61, 2014, doi:10.1007/s12596-013-0148-7.

[60] N. Cvejic, D. Bull, and N. Canagarajah, "Region-based multimodal image fusion using ica bases," *IEEE Sensors Journal*, vol. 7, no. 5, pp. 743–751, 2007, doi:10.1109/JSEN.2007.894926.

[61] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ica bases," *Information Fusion*, vol. 8, no. 2, pp. 131–142, 2007, Special Issue on Image Fusion: Advances in the State of the Art, doi:10.1016/j.inffus.2005.09.001.

[62] N. Mitianoudis, S.-A. Antonopoulos, and T. Stathaki, "Region-based ICA image fusion using textural information," in *2013 18th International Conference on Digital Signal Processing (DSP)*, pp. 1–6, 2013, doi:10.1109/ICDSP.2013.6622678.

[63] Z. Omar, N. Mitianoudis, and T. Stathaki, "Region-based image fusion using a combinatory chebyshev-ica method," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1213–1216, 2011, doi:10.1109/ICASSP.2011.5946628.

[64] J. Mou, W. Gao, and Z. Song, "Image fusion based on non-negative matrix factorization and infrared feature extraction," in *2013 6th International Congress on Image and Signal Processing*, vol. 2, pp. 1046–1050, 2013, doi:10.1109/CISP.2013.6745210.

[65] W. Kong, Y. Lei, Y. Lei, and J. Zhang, "Technique for image fusion based on non-subsampled contourlet transform domain improved NMF," *Science China Information Sciences*, vol. 53, pp. 2429–2440, 2010, doi:10.1007/s11432-010-4118-2.

[66] L. Li, Z. Xia, H. Han, G. He, F. Roli, and X. Feng, "Infrared and visible image fusion using a shallow CNN and structural similarity constraint," *IET Image Processing*, vol. 14, no. 14, pp. 3562–3571, 2020, doi:10.1049/iet-ipr.2020.0360.

[67] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 16, no. 03, p. 1850018, 2018.

[68] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: A unified unsupervised image fusion network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 502–518, 2020, doi:10.1109/TPAMI.2020.3012548.

[69] L. Tang, X. Xiang, H. Zhang, M. Gong, and J. Ma, "DIVFusion: Darkness-free infrared and visible image fusion," *Information Fusion*, vol. 91, pp. 477–493, 2023, doi:10.1016/j.inffus.2022.10.034.

[70] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Information Fusion*, vol. 48, pp. 11–26, 2019, doi:10.1016/j.inffus.2018.09.004.

[71] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614–2623, 2019, doi:10.1109/TIP.2018.2887342.

[72] H. Li, X.-J. Wu, and T. Durrani, "NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9645–9656, 2020, doi:10.1109/TIM.2020.3005230.

[73] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *ICLR*, 2021.

[74] H. Zhao and R. Nie, "DNDT: Infrared and visible image fusion via DenseNet and dual-transformer," in *2021 International Conference on Information Technology and Biomedical Engineering*, pp. 71–75, 2021, doi:10.1109/ICITBE54178.2021.00025.

[75] D. Rao, T. Xu, and X.-J. Wu, "TGFuse: An infrared and visible image fusion approach based on transformer and generative adversarial network," *IEEE Transactions on Image Processing*, pp. 1–1, 2023, doi:10.1109/TIP.2023.3273451.

[76] J. Li, J. Zhu, C. Li, X. Chen, and B. Yang, "CGTF: Convolution-guided transformer for infrared and visible image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022, doi:10.1109/TIM.2022.3175055.

[77] X. Zhang and Y. Demiris, "Visible and infrared image fusion using deep learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 10535–10554, 2023, doi:10.1109/TPAMI.2023.3261282.

[78] C. Sun, C. Zhang, and N. Xiong, "Infrared and visible image fusion techniques based on deep learning: A review," *Electronics*, vol. 9, no. 12, 2020, doi:10.3390/electronics9122162.

[79] X. Zhang, Y. Ma, F. Fan, Y. Zhang, and J. Huang, "Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition," *J. Opt. Soc. Am. A*, vol. 34, pp. 1400–1410, Aug 2017, doi:10.1364/JOSAA.34.001400.

[80] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infrared Physics & Technology*, vol. 82, pp. 8–17, 2017, doi:10.1016/j.infrared.2017.02.005.

[81] Z. Liu, Y. Feng, H. Chen, and L. Jiao, "A fusion algorithm for infrared and visible based on guided filtering and phase congruency in NSST domain," *Optics and Lasers in Engineering*, vol. 97, pp. 71–77, 2017, doi:10.1016/j.optlaseng.2017.05.007.

[82] Y. Yang, Y. Que, S. Huang, and P. Lin, "Multiple visual features measurement with gradient domain guided filtering for multisensor image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 4, pp. 691–703, 2017, doi:10.1109/TIM.2017.2658098.

[83] F. Meng, M. Song, B. Guo, R. Shi, and D. Shan, "Image fusion based on object region detection and non-subsampled contourlet transform," *Computers & Electrical Engineering*, vol. 62, pp. 375–383, 2017, doi:10.1016/j.compeleceng.2016.09.019.

[84] B. Zhang, X. Lu, H. Pei, and Y. Zhao, "A fusion algorithm for infrared and visible images based on saliency analysis and non-subsampled shearlet transform," *Infrared Physics & Technology*, vol. 73, pp. 286–297, 2015, doi:10.1016/j.infrared.2015.10.004.

[85] C. Liu, Y. Qi, and W. Ding, "Infrared and visible image fusion method based on saliency detection in sparse domain," *Infrared Physics & Technology*, vol. 83, pp. 94–102, 2017, doi:10.1016/j.infrared.2017.04.018.

[86] Z. Liu, H. Yin, B. Fang, and Y. Chai, "A novel fusion scheme for visible and infrared images based on compressive sensing," *Optics Communications*, vol. 335, pp. 168–177, 2015, doi:10.1016/j.optcom.2014.07.093.

[87] W. Kong, L. Zhang, and Y. Lei, "Novel fusion method for visible light and infrared images based on NSST–SF–PCNN," *Infrared Physics & Technology*, vol. 65, pp. 103–112, 2014, doi:10.1016/j.infrared.2014.04.003.

[88] C. Zhao, G. Shao, L. Ma, and X. Zhang, "Image fusion algorithm based on redundant-lifting NSWMDA and adaptive PCNN," *Optik*, vol. 125, no. 20, pp. 6247–6255, 2014, doi:10.1016/j.ijleo.2014.08.024.

[89] Z. Zhou, B. Wang, S. Li, and M. Dong, "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters," *Information Fusion*, vol. 30, pp. 15–26, 2016, doi:10.1016/j.inffus.2015.11.003.

[90] Z. Zhou, M. Dong, X. Xie, and Z. Gao, "Fusion of infrared and visible images for night-vision context enhancement," *Applied optics*, vol. 55, no. 23, pp. 6480–6490, 2016, doi:10.1364/AO.55.006480.

[91] Y. Ma, J. Chen, C. Chen, F. Fan, and J. Ma, "Infrared and visible image fusion using total variation model," *Neurocomputing*, vol. 202, pp. 12–19, 2016, doi:10.1016/j.neucom.2016.03.009.

[92] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Information Fusion*, vol. 31, pp. 100–109, 2016, doi:10.1016/j.inffus.2016.02.001.

[93] H. Guo, Y. Ma, X. Mei, and J. Ma, "Infrared and visible image fusion based on total variation and augmented lagrangian," *Journal of the Optical Society of America A*, vol. 34, no. 11, pp. 1961–1968, 2017, doi:10.1364/JOSAA.34.001961.

[94] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image and Vision Computing*, vol. 27, no. 10, pp. 1421–1432, 2009, Special Section: Computer Vision Methods for Ambient Intelligence, doi:10.1016/j.imavis.2007.12.002.

[95] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, "Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 94–109, 2012, doi:10.1109/TPAMI.2011.109.

[96] S. Paul, I. S. Sevcenco, and P. Agathoklis, "Multi-exposure and multi-focus image fusion in gradient domain," *Journal of Circuits, Systems and Computers*, vol. 25, no. 10, p. 1650123, 2016, doi:10.1142/S0218126616501231.

[97] S. Paul, "Multi-exposure and multi-focus image fusion in gradient domain," https://www.mathworks.com/matlabcentral/fileexchange/48782-multi-exposure-and-multi-focus-image-fusion-in-gradient-domain, 2016, [Online; Retrieved October 1, 2022].

[98] P. J. Hampton, P. Agathoklis, and C. Bradley, "A new wave-front reconstruction method for adaptive optics systems using wavelets," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 781–792, 2008, doi:10.1109/JSTSP.2008.2006386.

[99] Z. Guo, X. Yu, and Q. Du, "Infrared and visible image fusion based on saliency and fast guided filtering," *Infrared Physics & Technology*, vol. 123, p. 104178, 2022, doi:10.1016/j.infrared.2022.104178.

[100] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics*, vol. 21, p. 249–256, jul 2002, doi:10.1145/566654.566573.

[101] I. S. Sevcenco, P. J. Hampton, and P. Agathoklis, "A wavelet based method for image reconstruction from gradient data with applications," *Multidimensional Systems and Signal Processing*, vol. 26, no. 3, p. 717–737, 2015, doi:10.1007/s11045-013-0262-3.

[102] I. S. Sevcenco, P. J. Hampton, and P. Agathoklis, "Wavelet based image reconstruction from gradient data," https://www.mathworks.com/matlabcentral/fileexchange/48066-wavelet-based-image-reconstruction-from-gradient-data, 2015, [Online; Retrieved October 1, 2022].

[103] M. Vetterli and J. Kovačevic, *Wavelets and Subband Coding*, Prentice-Hall, Inc., 1995.

[104] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013, doi:10.1109/TPAMI.2012.213.

[105] K. He and J. Sun, "Fast guided filter," 2015, arXiv:1505.00996.

[106] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems*, pp. 474–485, Academic Press, 1994, doi:10.1016/B978-0-12-336156-1.50061-6.

[107] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *2018 24th International Conference on Pattern Recognition*, pp. 2705–2710, 2018, doi:10.1109/ICPR.2018.8546006.

[108] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations*, pp. 1–14, 2015.

[109] A. Toet, J. Ijspeert, A. Waxman, and M. Aguilar, "Fusion of visible and thermal imagery improves situational awareness," *Displays*, vol. 18, no. 2, pp. 85–95, 1997, doi:10.1016/S0141-9382(97)00014-0.

[110] A. Toet, "The TNO multiband image data collection," *Data in Brief*, vol. 15, pp. 249–251, 2017, doi:10.1016/j.dib.2017.09.038.

[111] Z. Zhou, M. Dong, X. Xie, and Z. Gao, "Fusion of infrared and visible images for night-vision context enhancement," *Applied Optics*, vol. 55, pp. 6480–6490, Aug 2016, doi:10.1364/AO.55.006480.

[112] H. Li and X.-J. Wu, "Infrared and visible image fusion using latent low-rank representation," 2018, 1804.08992, arXiv:1804.08992.

[113] H. Li, X.-J. Wu, and J. Kittler, "MDLatLRR: A novel decomposition method for infrared and visible image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4733–4746, 2020, doi:10.1109/TIP.2020.2975984.

[114] H. Li, X.-J. Wu, and T. Durrani, "Infrared and visible image fusion with resnet and zero-phase component analysis," *Infrared Physics & Technology*, vol. 102, p. 103039, 2019, doi:10.1016/j.infrared.2019.103039.

[115] L. Tang, J. Yuan, H. Zhang, X. Jiang, and J. Ma, "PIAFusion: A progressive infrared and visible image fusion network based on illumination aware," *Information Fusion*, vol. 83-84, pp. 79–92, 2022, doi:10.1016/j.inffus.2022.03.007.

[116] L. Tang, J. Yuan, and J. Ma, "Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network," *Information Fusion*, vol. 82, pp. 28–42, 2022, doi:10.1016/j.inffus.2021.12.004.

[117] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," *Computational visual media*, vol. 8, no. 3, pp. 331–368, 2022, doi:10.1007/s41095-022-0271-y.

[118] Digital Research Alliance of Canada, "Cedar," https://docs.alliancecan.ca/wiki/Cedar, 2019, [Online; Accessed March 1, 2023].

[119] X. Zhang, P. Ye, and G. Xiao, "VIFB: A visible and infrared image fusion benchmark," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 468–478, 2020, doi:10.1109/CVPRW50498.2020.00060.

[120] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'Information measure for performance of image fusion'," *Electronics Letters*, vol. 44, pp. 1066 – 1067, 02 2008, doi:10.1049/el:20081754.

[121] N. Cvejic, A. Loza, D. Bull, and N. Canagarajah, "A similarity metric for assessment of image fusion algorithms," *International Journal of Signal Processing*, vol. 2, no. 3, pp. 178–182, 2005.

[122] H. Chen and P. K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," *Information Fusion*, vol. 8, no. 2, pp. 193–207, 2007, Special Issue on Image Fusion: Advances in the State of the Art, doi:10.1016/j.inffus.2005.10.001.

[123] N. Cvejic, C. Canagarajah, and D. Bull, "Image fusion metric based on mutual information and tsallis entropy," *Electronics letters*, vol. 42, no. 11, p. 1, 2006, doi:10.1049/el:20060693.

[124] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Information Fusion*, vol. 14, no. 2, pp. 127–135, 2013, doi:10.1016/j.inffus.2011.08.002.