Fusion of Images and Videos using Multi-scale Transforms

Sarath Somasekharan Pillai

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science (Electrical & Computer Engineering)

at

Concordia University

Montréal, Québec, Canada

December 2013

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Sarath Somasekharan Pillai

Entitled: Fusion of Images and Videos using Multi-scale Transforms

and submitted in partial fulfilment of the requirements for the degree of

Master of Applied Science (Electrical & Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. M. Z. Kabir

_____ Examiner, External
Dr. T. Fancott (CSE) to the Program

_____ Examiner
Dr. M. O. Ahmad

_____ Supervisor
Dr. M. N. S. Swamy

Approved by _____
Dr. W. E. Lynch
Chair, Department of Electrical
and Computer Engineering

_____ 20 _____

Dr. C. W. Trueman
Interim Dean, Faculty of Engineering
and Computer Science

ABSTRACT

Fusion of Images and Videos using Multi-scale Transforms

Sarath Somasekharan Pillai

This thesis deals with methods for fusion of images as well as videos using multi-scale transforms. First, a novel image fusion algorithm based primarily on an improved multi-scale coefficient decomposition framework is proposed. The proposed framework uses a combination of non-subsampled contourlet and wavelet transforms for the initial multi-scale decompositions. The decomposed multi-scale coefficients are then fused twice using various local activity measures. Experimental results show that the proposed approach performs better or on par with the existing state-of-the art image fusion algorithms in terms of quantitative and qualitative performance. In addition, the proposed image fusion algorithm can produce high quality fused images even with a computationally inexpensive two-scale decomposition. Finally, we extend the proposed framework to formulate a novel video fusion algorithm for camouflaged target detection from infrared and visible sensor inputs. The proposed framework consists of a novel target identification method based on conventional thresholding techniques proposed by Otsu and Kapur et al. These thresholding techniques are further extended to formulate novel region-based fusion rules using local statistical measures. The proposed video fusion algorithm, when used in target highlighting mode, can further enhance the hidden target, making it much easier to localize the hidden camouflaged target. Experimental results show that the proposed video fusion algorithm performs much better than its counterparts in terms of quantitative and qualitative results as well as in terms of time complexity. The relative low complexity of the proposed video fusion algorithm makes it an ideal candidate for real-time video surveillance applications.

ACKNOWLEDGEMENTS

I would kindly like to express my deepest gratitude towards my thesis advisor, Professor M.N.S. Swamy for giving me this wonderful opportunity. It has been an immense honor and privilege for me to work under his supervision. I would also like to thank him for his constant encouragement, guidance and financial support throughout the course of this thesis.

I had a very wonderful time in the signal processing lab EV. 10.119 for the past 16 months. For which, I would like to thank my colleagues Mufleh Al-Shatnawi, Ali Baghaki and Yaser Mohammad Taheri.

I am especially grateful to Paul Leons for his support and for being a great friend. Initially, life in Montreal was hard, but was made easy for which I owe a million thanks to Mohan Kumar Vengatachalam, Srinivasan Rajivelu, Hariharan Guhan Satish Tamilselvan, Selva Ganesh Elangovan, Ankit C. Mehta, Denny Chacko Jacob, Ganghadharan Kalyanasundharam, Selva Kumar Raju, Rahul Paul, Praveen Kumar Korrai, Shreyamshakumar, Gilbert Makita and Kumar Mani. Without them, life in Montreal would not have been the same.

To my parents

&

To jalaja appachi

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$Q_{AB/F}$	Edge based similarity measure
E	Entropy
A_K	Approximate wavelet coefficient
$M(L)$	Match measure
$T(L)$	Adaptive threshold
L_k	NSCT low frequency coefficients
$B_k(g,m)$	NSCT bandpass coefficients
$D(i)$	Wavelet detailed coefficients
ϕ	Weight smoothing function
$E(L_c)$	Local energy calculated within a window
$w(m,n)$	Local window size
L_w	Low frequency weighted averaging schemes
L_s	Low frequency selection schemes
F_L	Fused NSCT low frequency coefficients
F_B	Fused NSCT bandpass coefficients
$F_D(i)$	Fused wavelet detailed coefficients
$F_{initial}$	Fused initial image or image frame
F_{final}	Fused final image or image frame
H_k	Kapur threshold
T_h	Otsu threshold
$\Psi_S(i,j)$	Sum-modified laplacian based activity measure
$\phi_B(i,j)$	Activity map based on Otsu threshold

LIST OF ACRONYMS

SNR	Signal to noise ratio
MRA	Multi resolution analysis
MST	Multi scale transforms
MRI	Magnetic resonance imaging
LP	Laplacian pyramid
DWT	Discrete wavelet transform
NSCT	Non-subsmapled contourlet transform
HOSVD	Higher order singular value decompsoition
CT	Computed tomography
PET	Positron emission tomography
SPECT	Single photon emission tomography
MGA	Multi-scale geometric analysis tools
MSD	Multi-scale decomposition
SF	Spatial frequency
PC	Phase congruency
PCNN	Pulse coupled neural networks
LSD	Local standard deviation
SML	Sum-modified laplacian
CVT	Curvelet transform
DTCWT	Dual-tree compex wavelet transform
CT	Contourlet transform
NSPFB	Non-subsampled pyramidal filter banks
NSDFB	Non-subsampled directional filter banks

Chapter 1

Introduction

1.1 Background

The revolutionary development in the field of image sensors and image acquisition techniques has led to the availability of a significant amount of information from an observed scene. These significant features can be extracted by using various sensory modalities such as lidar, radar, sonar, infrared and thermal cameras. The output images produced by these sensors will have contrasting texture properties and salient features. Nowadays, numerous areas of experimental sciences, namely, video surveillance, satellite imaging, remote sensing and medical imaging require a unique combination of these multi-modal information. This can be primarily achieved by devising a computerized scheme of blending critical information from various sensor inputs to a more comprehensive composite fused image. This automatic way of merging valuable multi-sensor data is known as image or video fusion.

The paramount encouragement for images or video fusion is to enhance the overall quality of the information preserved in the final fused output. Numerous studies of the conventional image or video fusion methodologies have validated their benefits for various applications [1, 2] and some of them are as follows:

1. Increased dynamic range of operation.

2. Increased spatial content.

3. Increased system reliability.

4. Compressed portrayal of information.

The multiple input images of an image fusion system can be obtained either with a single camera at multiple instants or by various cameras at a particular instant. For example, consider the case of camouflaged target detection (Figure 1.1) system consisting of sensors sensitive to multiple wavebands, including color visible and thermal infrared cameras. Typically, these systems will have a unique screen display which is only capable of showing data from an individual image sensor, thereby resulting in an information overload, as the operator has to cycle through the multiple sensor outputs at a given instant of time. These sensor outputs can be effectively combined together into a single and superior image or image frame representation by using a suitable image or video fusion framework.

A **B** **C**

Figure 1.1: Camouflaged target detection: Source images [3]: Thermal image frame (A), Visible image frame (B), Fused image: Thermal-visible fused image (C)

1.1.1 Basic image or video fusion system

The basic schematic of an image or video fusion system is shown in Figure 2.4. The various blocks involved in the schematic are described below.

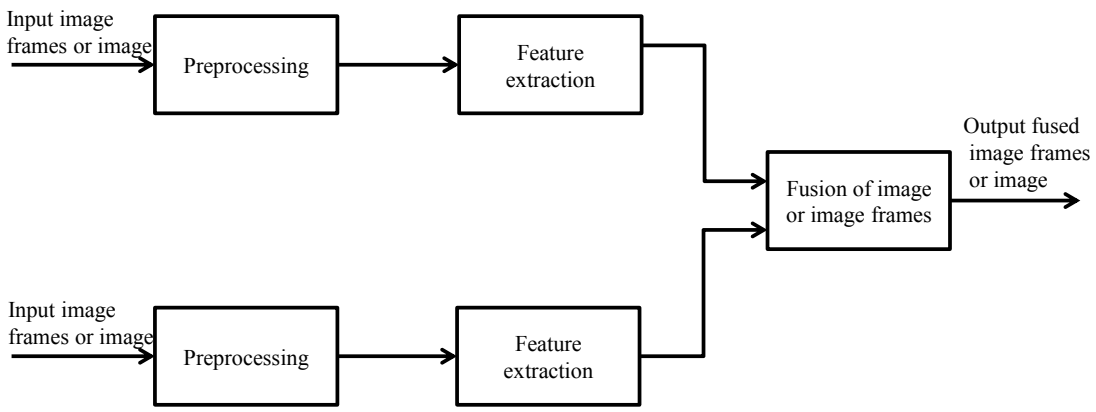

Figure 1.2: Block diagram of a basic image or video fusion system

1. Preprocessing

   Depending upon the nature of data acquisition sensors, various operations like contrast enhancement, registration and denoising are performed.

2. Feature extraction

   In this module, various characteristic or salient features from the sensor inputs are extracted using certain activity measures. These measures are designed in accordance with the nature of the imaging modality. The features obtained are then used to form a decision map.

3. Fusion of image or image frames

   Here, the salient features of the image frames or images are combined by using suitable fusion rules to form the output fused image.

1.1.2 Classifications of image or video fusion algorithms

In general, image or video fusion techniques can be implemented at the pixel-level, feature-level, or the decision-level [4]. Among them, the pixel-level techniques are widely preferred for real-time applications as they are computationally efficient and offer true measured quantities at the lowest level. These pixel-level fusion techniques can be further classified as below:

1. Image or video fusion based on pixel-averaging

   Pixel averaging or additive fusion approach is one of the simplest image or video fusion techniques. In this approach, a weighting function is applied to each of the input images or image frames and is then fused by taking a linear combination among the various inputs. In other words, X % of one input image or image frame is linearly fused with $100 - X$ % of the other input image or image frame. This additive combination results in a low-latency implementation consuming relatively less processing power. The fused output is clearly an additive combination of the various input images or image frames and is preferable than having two camera outputs side-by-side. However, in some cases like camouflaged target detection, a linearly weighted fusion algorithm will result in an image or image frame with variable quality fused output. This will result in the loss of significant scene features such as texture and thermal information, and the fused image or image frame might not produce an enhanced view of the hidden target. Furthermore, in the case of a feature appearing in any of the two sensor inputs, averaging will result in the loss of the feature contrast, or may not be contained in the output at all.

2. Image or video fusion based on multi-resolution analysis

   An ideal image or video fusion output should have a high signal to noise ratio (SNR) value. This can be achieved by using more advanced techniques such as multi-resolution analysis (MRA) at the pixel-level. The main advantages of using MRA techniques over conventional pixel-averaging method are as follows.

   (a) The significant features present in the inputs are extracted at various resolutions. This multi-level extraction of different scale results in the transfer of maximum scene content from the input to the output fused image or image frame.

(b) The features of sensor inputs will be represented completely even if these features are present in any of the two sensor inputs.

(c) The output fused image is free from contrast reduction and artifacts and can be further enhanced by incorporating suitable weighting functions.

Among the MRA techniques, the most commonly used ones are multi-scale transforms (MST) [5, 6]. These MST-based image or video fusion algorithms primarily decompose images or image frames into geometric representations having a different scale of resolutions and can be classified as follows:

(a) Lowest resolution stage

The low-frequency components from each sensor input are contained in the lowest resolution stage. The components contained at the lowest resolution stage are usually a blurred version of the input images or image frames.

(b) Mid-resolution stage

The information about features having different sizes is contained in the mid-resolution representations.

(c) Highest resolution stage

The sharp details of the input images or image frames like edge details are contained in the highest resolution stage.

Some of the most commonly-used multi-scale transforms in the literature for image or video fusion are the Laplacian pyramid [7]; the discrete wavelet transform (DWT) [8] and the non-subsampled contourlet transform (NSCT)[9].

1.1.3 Characteristic features of a fusion algorithm

Some of the most prominent or characteristic features of a fusion algorithm are described below.

1. Feature preservation

   The output fused image or image frame should conserve all the redundant and complementary features from the individual sensor inputs.

2. Immunity from artificats

   The final fused or composite image or image frame should be free from spurious artifacts. In other words, the fusion rules should not cause any distortion or inconsistency to the final fused image or image frame.

3. Shift or rotational independence

   The quality of the final fused image or image frame should not be highly sensitive or variant towards the location or orientation of various objects present in the individual sensor inputs.

1.1.4 Potential applications of fusion algorithms

1. Medical imaging

   In the case of medical imaging, usually it is very difficult for a physician to diagnose diseases like tumors from an individual imaging modality. In this scenario, image fusion can be used to assist the physicians by extracting and combining features from different imaging modalities, which cannot be observed by using a single imaging modality [10]. The various imaging modalities such as magnetic resonance image (MRI), computed tomography (CT), single photon emission computed tomography (SPECT), positron emission tomography (PET), and ultrasound are characterized by their own peculiar properties. For example, MRI imaging modalities (MRI-T1 and MRI-T2) produce an output having a significant amount of information about various anatomical structures and greater contrast between the various normal and abnormal body tissues. Modalities such as SPECT, PET, and CT are enriched with significant information highlighting the functional and metabolic activities. Currently, the

MRI modalities are often registered with imaging modalities such as SPECT, PET and CT to obtain an exact idea about both the anatomical and functional details. This results in an information overload, as the physician has to cycle through multiple outputs for a proper diagnosis. Thus, image fusion is performed to extract and combine maximum complementary and contrasting information into a single output image. This will result in a better and accurate diagnosis as well as lower the storage space due to the substantial minimization in file size.

2. Optimization of dynamic range and depth of field

   The dynamic range of each sensor has a certain limit. Usually, the highlights are washed out when the exposure is increased to enhance the details of a shadow in a scene. Similarly, the details of the shadow are lost when bright parts of the image are exposed. The dynamic range of these images can be tremendously enhanced by fusing images obtained with distinctive sensor gains or unlike apertures. The depth of field of all optical imaging systems is limited. The depth of field or scene sharpness of a foreground with respect to a background can be enhanced by limiting the aperture of the lens. However, at the same time it results in the loss of sensitivity and the resolution of the camera. This can be avoided by increasing the effective depth of the field involved by combining images from different cameras having foci at diverse distances from the lens of the camera.

3. Detection of camouflaged targets

   The targets that are hidden from human or machine by changing its appearance similar to the background is known as camouflaged targets. These targets cannot be effortlessly distinguished by using inputs from single waveband based cameras. For example, in the infrared band, a target can be readily identified but all its background information like edge, color and contrast will

be lost. In the case of a visible camera, all the background information will be retained but target information will be completely lost. The target in the input video streams can be better identified and detected by fusing these inputs into a single output video stream. The fused output video will no longer contain a camouflaged target as it retains the properties of both thermal and visible input streams.

4. Monitoring adverse climatic conditions

   In adverse weather conditions, the performance of imaging systems degrades quite rapidly and is not reliable. Conventional imaging systems will result in substandard visible contrast or relatively shallow thermal contrast with the presence of smoke or foggy weather conditions. These systems can be made more reliable in adverse weather conditions by fusing the video streams obtained from the different wavebands.

5. Real-time surveillance of airborne targets

   The tracking and searching of ground based targets in military and paramilitary applications is usually carried out by using thermal and visible cameras. By using a thermal camera, the targets can be freely tracked, and their positions can be absolutely located. On the other hand, the ground features are tracked by using a visible color camera. By the fusion of the data from both cameras, the targets can be easily localized and tracked with respect to the corresponding background retained from the visible camera.

6. Real-time surveillance for ground based security systems

   In many public security checkpoints like airports, scanners based on multiple wavebands are used for inspecting people and their luggage. These scanners have ranges spanning from ultraviolet wavebands to terahertz wavebands. The objects that are hidden can be easily detected and identified by combining multiple waveband inputs.

7. Driver aids in automotive applications

   To improve the visibility of the driver as well as for the safety of pedestrians, numerous driver aids are being developed. Normally, these driver aids have multiple accompanying imagery displays, which distracts the driver. This distraction can be avoided by fusing the consequent imagery into a single display, thereby also improving the driver's awareness of the situation.

1.2 Motivation

In general, image or video fusion techniques vary from pixel averaging to transform domain approaches such as Laplacian pyramid [7] and higher order singular value decomposition (HOSVD) [2]. Among them, the most commonly used ones are the transform domain-based multi-scale transforms, which decompose a signal into various scales and directional subbands. In recent times, they have become a lot more popular with the availability of a group of multi-scale transforms known as multi-scale geometric analysis (MGA) tools [6]. These tools are primarily based on a mechanism similar to that of a human visual system. As a result, the outputs produced by the image fusion algorithms incorporating these MGA tools are more visually pleasing and are free from traditional drawbacks such as artifacts and reduced contrast.

Non-subsampled contourlet transform (NSCT) [9] is one such MGA tool. It is a highly effective shift-invariant multi-scale transform and its effectiveness for various multi-modal image fusion applications has been further established on the basis of a performance study conducted by Li et al. [5]. The existing NSCT methodology [11] available in the literature is computationally more demanding. This is mainly due to the fact that they use a configuration of four scales and directional subbands of the order of 4, 8, 8 and 16 for multi-scale decomposition (MSD). This leads to

a multi-scale representation of a very large dimension resulting in a high computational complexity. The authors in [12] have tried to address this issue to an extent by combining NSCT with a Haar wavelet transform for MSD. The decomposed coefficients are then fused by using computationally complex activity measures such as phase congruency (PC) and spatial frequency (SF)-motivated pulse coupled neural networks (PCNN). As a result, the aforementioned framework also suffers from high computational complexity.

1.3 Objective of the thesis

The main theme of this thesis is to develop two schemes for the fusion of images and videos obtained from different multi-modal sensors. The algorithms suggested in this thesis are based on a two-stage fusion architecture. This framework primarily uses a combination of multi-scale transforms, and is obtained by a combination of wavelets and non-subsampled contourlets. Initially, this hybrid multi-scale transform is coupled with novel pixel-based fusion rules for the fusion of images from multi-modal sensors. These fusion rules employ local statistical measures as activity measures for recognizing and transferring salient features from the individual sensor inputs. As a result, the proposed image fusion algorithm is highly successful in inducing fused images of superior quality even with a computationally inexpensive two-scale decomposition. Finally, the above-mentioned framework is further incorporated by region-based fusion rules for the detection of camouflaged targets. Here, the rules used for fusion are primarily based on the thresholds given by Otsu [13] and Kapur et al. [14]. These thresholds detect and highlight the hidden target in the fused output video.

1.4 Organization of the thesis

The thesis is organized as follows. In Chapter 2, we give the relevant background knowledge associated with multi-scale transform oriented image fusion. Initially, their fundamental aspects are explained, followed by a discussion of the relevant aspects of two widely-used multi-scale transforms, namely, DWT and NSCT. We then describe the advantages of using a combination of DWT and NSCT rather than a conventional multi-scale transform for image fusion.

In Chapter 3, we explain the fusion of images using pixel-based image fusion rules. A brief review related to the state-of-the-art pixel-based image fusion algorithms is also included. We describe a new multi-scale transform-based image fusion framework based upon a combination of wavelets and contourlets. Then, we propose a new pixel based image fusion algorithm using the above-mentioned framework and regional activity measures such as local-standard deviation (LSD) and sum-modified Laplacian (SML) [15]. The performance of the proposed algorithm is then compared with numerous state-of-the-art pixel-based approaches by applying frequently-employed qualitative parameters.

In Chapter 4, we present our contribution towards real-time video-surveillance using the above-mentioned novel multi-scale framework and region-based fusion rules. We then consider the use of our suggested algorithm towards camouflaged target detection for exposing hidden targets. Chapter 5 contains the conclusions along with some future directions for research on the topic considered in the thesis.

Chapter 2

Multi-scale domain based image fusion

In this chapter, various background material related to this thesis are discussed. Initially, the general concepts involved behind multi-scale transform-based image fusion are discussed, followed by a detailed description of the two most highly effective multi-scale transforms used for image fusion, namely, discrete wavelet transform and non-subsampled contourlet transform. Finally, this chapter concludes with a detailed description of the metrics used for the performance evaluation of the image fusion algorithms.

2.1 General architecture for multi-scale transform based fusion system

Image fusion techniques are mainly preferred in the transform domain rather than in the pixel domain. This is mainly for achieving a better representation of the significant features of the input imaging sensor. The most widely-used transform domain techniques are mainly based on multi-scale transforms. These transforms are

capable of representing the various significant features of the input sensory module at different scales, similar to the mechanism of the human visual system. The basis functions of these multi-scale transforms behave in a manner similar in terms the selection and identification of regional features like edges and lines. The two main components involved in a multi-scale transform-based fusion system are multi-scale transforms and fusion blocks. The basic architecture of a multi-scale transform-based fusion system is shown in Figure 2.1.

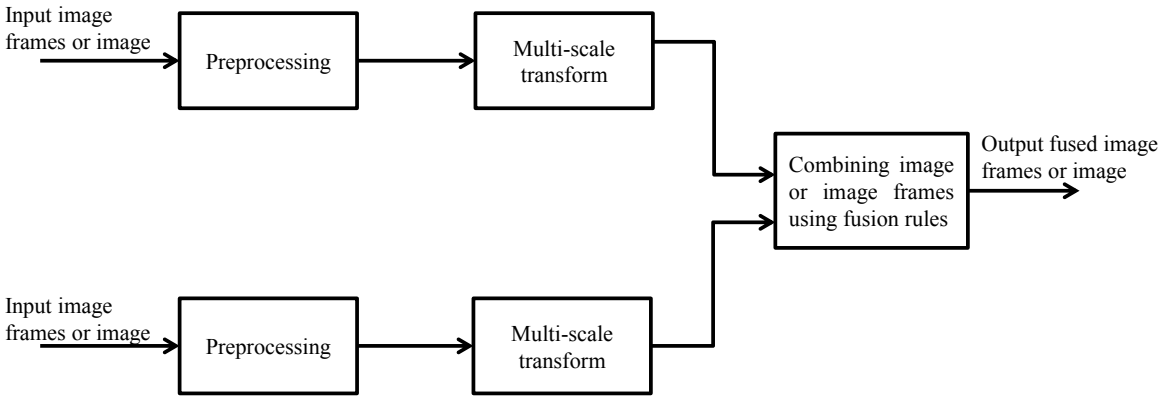

Figure 2.1: Block diagram of a multi-scale transform based fusion system

2.1.1 Multi-scale transforms

Multi-scale transforms in general decompose an imaging sensor input into multi-scale coefficients having different angles of orientation. The scale of decomposition and the direction of orientation are adjusted mainly in accordance with the resolution and nature of the sensor input. Numerous multi-scale transforms are available in the literature, such as the Laplacian pyramid transforms, the discrete wavelet transform (DWT), the curvelet transform (CVT), the dual-tree complex wavelet transform (DTCWT), the contourlet transform (CT) and the non-subsampled contourlet transform (NSCT). Among them, the DWT and NSCT offer numerous advantages and hence are more widely used than their counterparts for real-time fusion systems.

Discrete wavelet transforms

Discrete wavelet transform symbolizes the discrete sampling of the input signal. The various factors for the process of dilation and translation will be decided in such a way that it results in dyadic sampling. For a given scale N, a definite amount of translation is employed. This will result in a definite number of scaling and wavelet components.

$$f(m) = \sum_{R} C_{NR}\Phi_{NR} + \sum_{n=1}^{N}\sum_{R} D_{nR}\psi_{nR}(m) \tag{2.1}$$

where C_{NR} and D_{nR} correspond to the scaling and wavelet coefficients.

The principal term in the above-mentioned expression produces the approximate component of the signal. The detailed information present at the various scales from the origin to the existing resolution N is produced by the last term in (2.1). For a single scale decomposition, these discrete wavelet transforms will result in low and high frequency components. These low frequency components can be further decomposed to obtain the desired scale of resolution. Normally, single scale discrete wavelet transform decomposition is enough for the purpose of image fusion. The existing discrete wavelet transform has mainly two types of configurations. They are

- Decimated discrete wavelet transform
- Undecimated discrete wavelet transform

Among these two configurations, in our proposed image fusion framework we have used the decimated discrete wavelet transform. So, further discussion will be only on the decimated configuration of the discrete wavelet transform.

Decimated Discrete Wavelet Transform

In the decimated configuration of the discrete wavelet transform, the signal will be further down-sampled after each particular level of multi-scale transformation. In

the case of a two dimensional image, one among every two rows and columns will be retained after each level of decomposition. Therefore, each successive layer will be half of the size of the previous layer. This results in a structure similar to that of a pyramid having coarser details at the bottom.

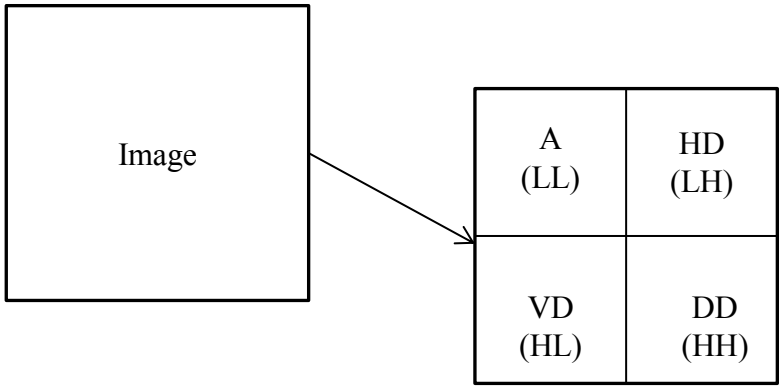

Figure 2.2: Schematic Diagram of a single scale DWT decomposition ([16])

When decimated discrete wavelet transform is applied to an image for multi-resolution analysis, it actually involves a two stage process. Initially, the techniques of filtering and down-sampling applied to the rows. After this step, the process of filtering and down-sampling are again repeated with the columns. This will ultimately result in four image layers of lower resolutions. Among them, one will be the approximate component and remaining will be the high frequency components such as vertical, horizontal and diagonal components.

The decimated version of the wavelet transform when used with more than one level of decomposition is highly susceptible to registration issues. This is mainly due to shift-variant nature of the decimated discrete wavelet transforms. In addition, the feature which does not have a horizontal or vertical orientation will be highly suppressed by the decimation process involved in the decimated discrete wavelet

transform. So, discrete wavelet transforms when used for image fusion with multi-levels of decomposition usually result in the introduction of artifacts in the fused output.

Non-subsampled contourlet transform

Non-subsampled contourlet transform [9] is a shift-invariant multi-scale transform which is derived primarily from the contourlet transform using non-subsampled filter banks. Among these filter banks [9], non-subsampled pyramidal filter banks (NSPFB) are responsible for the various levels or scales of decomposition, whereas non-subsampled directional filter banks (NSDFB) decompose these levels into various directional components.

The non-subsampled pyramidal filter banks are two-channel filter banks, which are primarily responsible for the multi-scale property of the NSCT. During each stage of the pyramidal decomposition, one low and one high frequency components are produced. This low frequency component on further decomposition by non-subsampled pyramids facilitates the encapsulation of various singularities present in the image. Ultimately, this mode of decomposition will result in N+1 subbands, out of which one will be a low frequency component and the remaining N high frequency components. The size of these N subbands will be the same as that of the input image. The value of N will be equal to the number of decomposition levels or scales.

A non-subsampled directional filter bank is also a two-channel filter bank, which is obtained by merging the various directional fan filter banks. These filter banks will result in a directional decomposition of the high frequency subbands obtained from the non-subsampled pyramidal filter banks. Normally, these filters result in a K-level decomposition and will produce 2^K directional subbands. The size of each subband will be the same as that of the source image. In addition, this multi-directional decomposition results in highly accurate directional information.

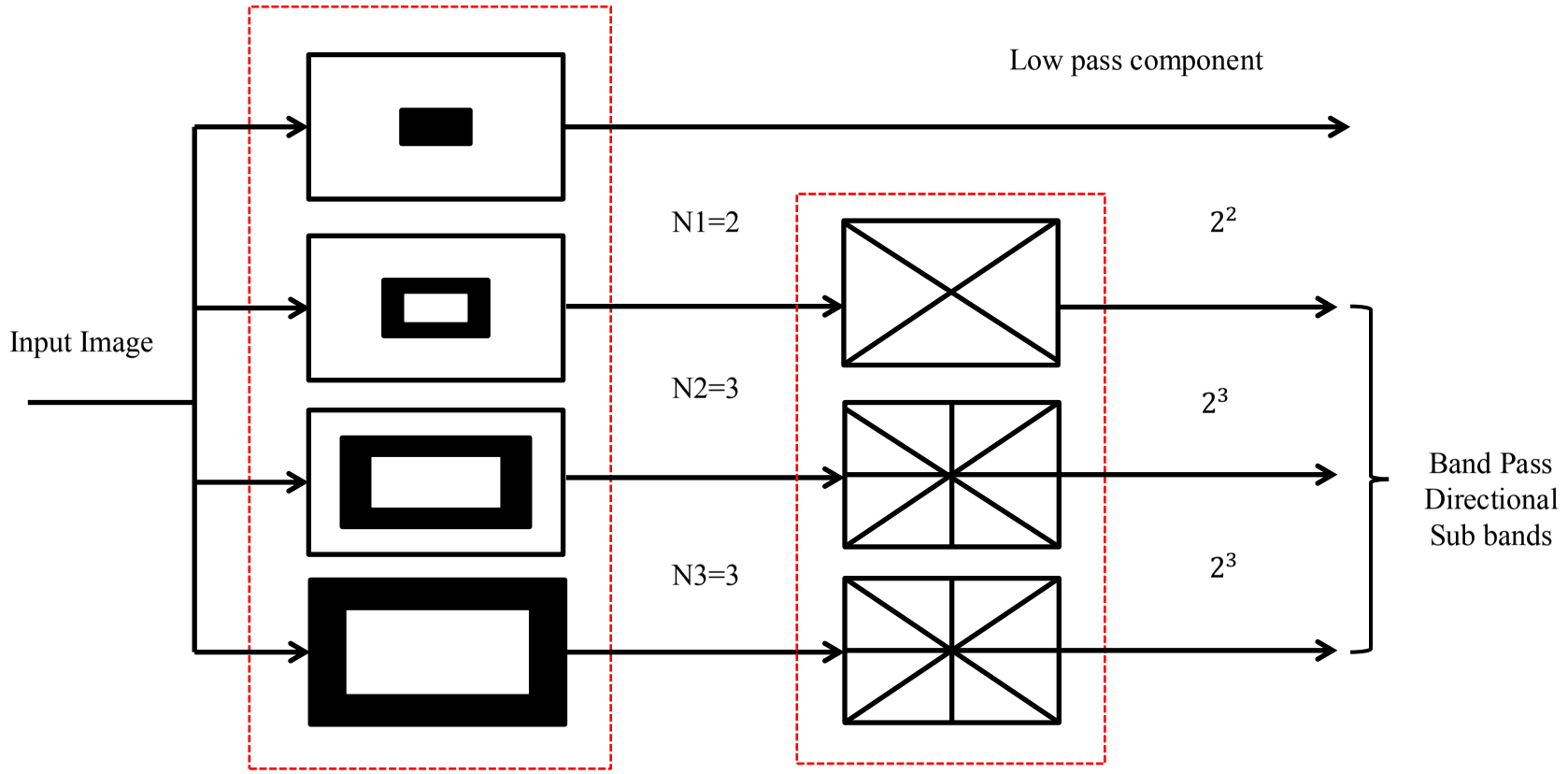

Figure 2.3: NSCT structure for a 3 level decomposition having directional subbands of the order of 4, 8, 8. [9]

.

Advantages of using a combination of DWT and NSCT

From the above discussion, it is clear that both the discrete wavelet transform and the non-subsampled contourlet transform have their own advantages and disadvantages. In our proposed framework, we have used a hybrid multi-scale transform which is formed by combing the DWT and the NSCT. This combination results in numerous advantages compared to the existing implementations with the DWT or the NSCT and are as follows.

1. Due to the coupling with the decimated discrete wavelet transform, the size of the input image will be reduced exactly by one half before the NSCT decomposition. As a result, the resulting hybrid multi-scale transform is computationally light compared to that of the conventional NSCT.

2. This unique combination is highly immune towards registration issues, whereas the DWT is easily prone to registration issues.

3. The fused output produced by this combination is free from fused artifacts, whereas outputs produced by the DWT are normally distorted with fused artifacts.

4. The hybrid combination can also produce high-quality fused outputs even with a two-scale decomposition. In the same scenario, conventional NSCT requires at least 4 scales of decomposition.

2.1.2 Classification of fusion rules

In general, fusion rules are basically classified into two classes. They are

- Pixel-based fusion rules.

- Region-based fusion rules.

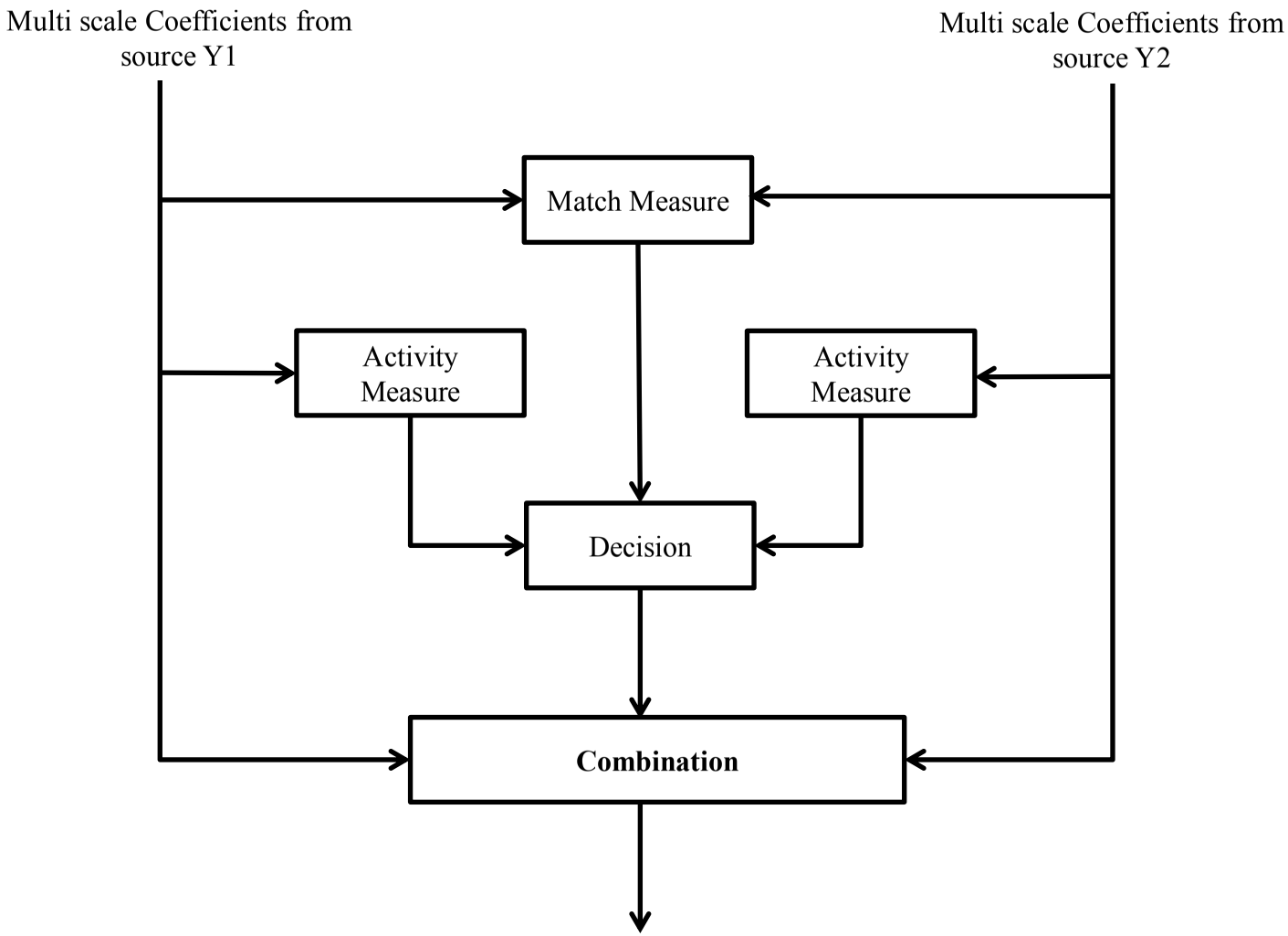

Figure 2.4: Schematic Diagram of a basic pixel-based fusion architecture [16]

Generic pixel-based fusion architecture

The basic components of a generic pixel-based system are as follows.

1. Activity measure

   This module is responsible for assigning an activity level of each pixel of the input image, which varies in accordance with its characteristic traits such as local statistics. These local statistics may vary from an absolute value of a single pixel to the local energy of a pixel neighborhood. In the case of a local neighborhood, the activity measure will indicate the presence of a desired feature if the corresponding average local energy or local standard deviation is high.

2. Match measure

   This unit is primarily accountable for indicating the extent of similarity between the input images at the various corresponding pixel locations. The simplest form of calculating the match between two inputs is to analyze the relative amplitudes of the inputs. But, a more robust match measure is mainly based on the average local correlation of the input images over a neighborhood of pixels. A low value of match measure indicates that the inputs are clearly different from each other at that particular pixel location. The similarity of the input images at a particular pixel location will be indicated by a very high value of the match measure.

3. Decision block

   This component chooses the tangible combination of the multi-scale coefficients from the input images. The decisions are primarily made on the basis of the inputs from the activity and match measures. These decisions can also be tuned up to consider various other dependencies such as spatial, inter and intra-scale dependencies.

4. Combination

   This element describes the actual merging of the transform coefficients. This combination can be either in the form of a linear or non-linear mapping.

Generic region-based fusion architecture

In this architecture, instead of fusing pixels one by one, they will be fused on the basis of regions. The region-based fusion system is almost similar to that of the generic pixel-based image fusion system except for a segmentation block. This segmentation block is introduced for partitioning the coefficients at the various scales and also in guiding various other fusion blocks.
The general segmentation module developed by Piella involves the following process [16].

1. Initialization

   In this step, a sampling scheme similar to that involved in multi-scale transform is applied. In addition, greater importance is placed on ensuring that the approximation pyramid obtained is completely dissimilar from that of the multi-scale one.

2. Linking

   The various relationships such as the child-parent relationships at the neighboring scales are established. This is largely done on the basis of the peculiar geometrical properties of the coefficients.

3. Root labeling

   Among the various input samples, those samples having relatively feeble connection with their parent coefficients and those in the highest level of decomposition are marked as roots.

4. Downward projection

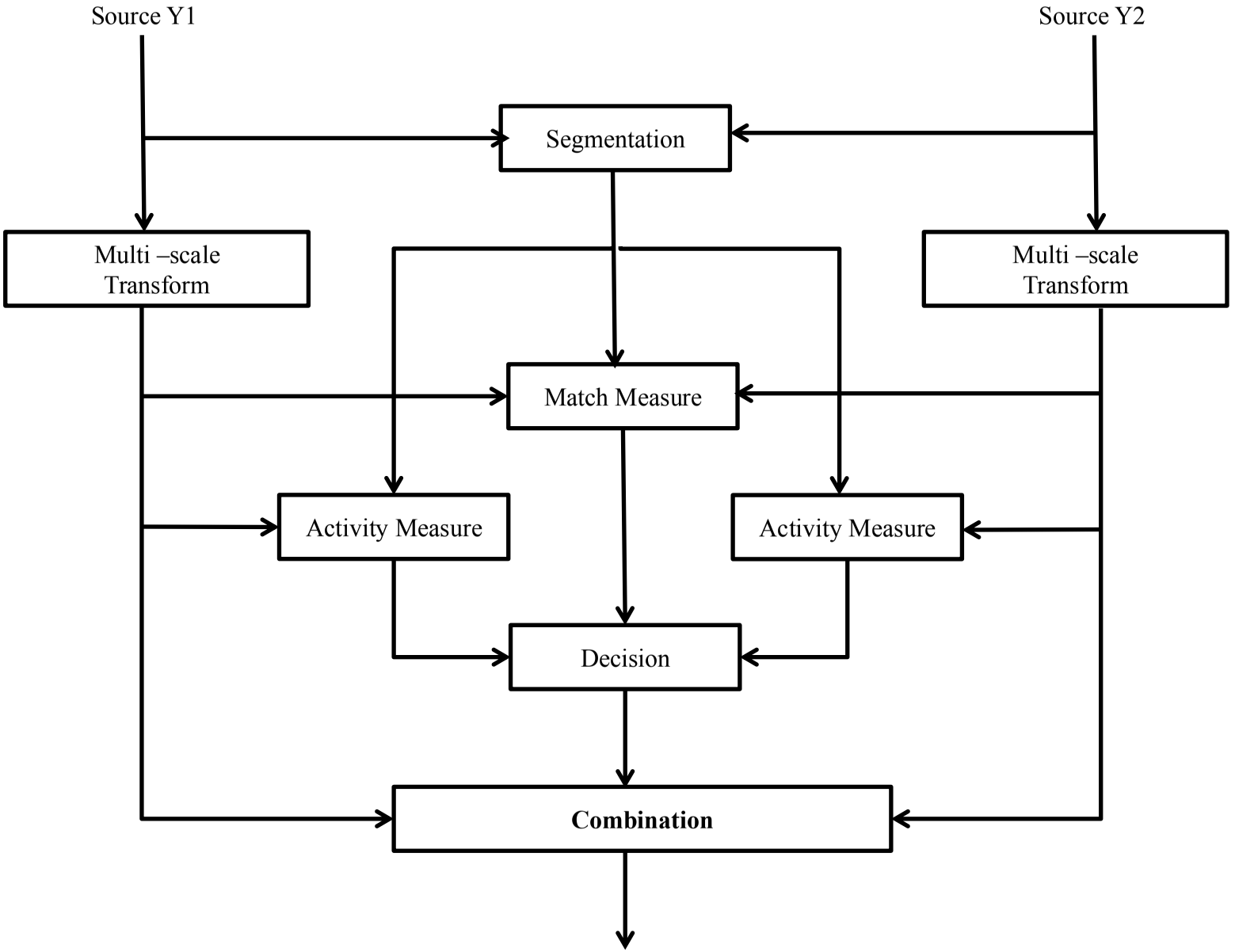

Figure 2.5: Schematic Diagram of a basic region based fusion architecture [16]

The image segmentation is basically carried out by sketching back the connections between the roots and the bottommost levels.

2.2 Quantitative metrics for the performance evaluation of image fusion schemes

For a quantitative analysis of the results, the most commonly used image fusion performance measures employed are the $Q_{AB/F}$ metric [17] and the entropy.

1. $Q^{AB/F}$ metric (edge-based similarity measure): $Q^{AB/F}$ metric [17] reflects the total amount of edge information preserved in the output fused image. The edge-based similarity metric is defined as follows:

$$Q^{AB/F} = \frac{\sum_{k=1}^{P}\sum_{h=1}^{Q}\left[Q_{k,h}^{AF}D_{k,h}^{x} + Q_{k,h}^{BF}D_{k,h}^{y}\right]}{\sum_{k=1}^{P}\sum_{h=1}^{Q}\left[D_{k,h}^{x} + D_{k,h}^{y}\right]} \tag{2.2}$$

   where A and B are the inputs and F is the output fused images. The terms $Q_{k,h}^{AF}$ and $Q_{k,h}^{BF}$ corresponds to the edge strength of the input images A and B. The functions $D_{k,h}^{x}$ and $D_{k,h}^{y}$ represents the weight function for the inputs A and B at a particular location x, y. The value of the $Q_{AB/F}$ metric varies between 0 to 1. The quality of the output fused image will be better, if the value of the $Q^{AB/F}$ metric is high.

2. Entropy:

   This quantitative measure reflects the total amount of information present in the output fused image. The entropy E of an image is calculated using

$$E = -\sum_{i=0}^{L-1} P_i log_2 P_i \tag{2.3}$$

   where L indicates the number of the gray levels involved and P_i denotes the ratio of the number of pixels having a gray level value i to the total number of

pixels present in the image. A higher entropy value suggests the effectiveness of the image fusion algorithm in preserving the contrast of an image.

2.3 Summary

In this chapter, we have discussed the basics of two most commonly-used multi-scale transforms and the advantages of combining the discrete wavelet and the non-subsampled contourlet transforms for multi-scale decomposition. In addition, we have also discussed about the basic architecture of region-based and pixel-based fusion rules. Finally, we have considered the commonly-used qualitative metrics for evaluating a given image fusion scheme.

Chapter 3

Low complexity image fusion algorithm using multi-scale transforms

In the previous chapters, we have briefly discussed some preliminary concepts regarding image fusion techniques and multi-scale transforms. The remaining part of the thesis is based on the above-mentioned fundamental concepts. In this chapter, we propose a novel low complexity image fusion algorithm using the multi-scale transforms [18]. The performance of the proposed algorithm is evaluated both from quantitative and qualitative points of view and compared with that of the other existing methods. The execution time of the proposed algorithm is also compared with that of their conventional counterparts.

3.1 Introduction

Over the last two decades, the field of multi-modal image sensors has been galvanized by the significant evolution of the semiconductor industry. Nowadays, these multi-modal sensors are used in a wide variety of fields ranging from remote sensing

to medical imaging. The principle idea behind the deployment of these multi-modal sensors is to extract maximum information contained in a scene. In short, image fusion can be defined as a technique or a process to obtain a composite representation or a visually enhanced scene by integrating various redundant and complementary information depicted in the same scene.

Image fusion techniques are generally employed in the pixel-based level due to their computational simplicity. These techniques vary from simple pixel averaging methods to more robust approaches such as multi-resolution analysis (MRA) using multi-scale transforms (MST). One of the pioneering works in the field of pixel-based image fusion techniques is that of the Laplacian pyramid (LP) [7]. The principle idea behind this approach is to decompose the coefficients using a Laplacian pyramid and fusing the coefficients using suitable fusion rules. The fused output produced by this approach suffers from the issues of reduced contrast and artifacts. In spite of all these drawbacks, this approach involves a very low computational complexity and is still considered as a benchmark for the pixel-based fusion methods in terms of the execution time. Recently, Liang et al. [2] proposed an image fusion algorithm based on the higher order singular value decomposition (HOSVD), which is primarily based on the ability of HOSVD to extract features of the high dimensional data. This method is fairly successful in preserving contrast of the fused image; however, it suffers from high computational complexity. Another type of image fusion approach which has recently attracted a lot of interest is based on the non-subsampled contourlet transform (NSCT) [9]. This is mainly due to its suitability towards fusion of various multi-modal inputs [5]. Numerous NSCT based image fusion algorithms are available in the literature. Most of them involve the use of NSCT for multi-scale decomposition and then combining inputs using suitable fusion rules. Among them, the most successful one in terms of the performance is the one proposed by Qu et al [11]. This algorithm primarily uses NSCT for a four scale decomposition and then

fuses the decomposed coefficients using spatial frequency (SF)-motivated pulse coupled neural networks (PCNN). This algorithm is also a highly complex one, mainly due to the usage of a four scale NSCT decomposition and also due to the usage of computationally expensive fusion rules based on PCNN.

To address the above mentioned limitations, we propose a novel image fusion algorithm, which is primarily based on an improved multi-scale coefficient decomposition framework. The main contribution of the proposed framework is the overall reduction in the number of scales of decomposition. These decomposed multi-scale coefficients are then fused by fusion rules based on computationally inexpensive local activity measures.

3.2 Proposed pixel-based image fusion algorithm

The architecture of the proposed image fusion algorithm is as shown in Fig. 3.1. The key idea behind the proposed algorithm is an improved coefficient decomposition framework using the multi-scale transforms. This framework is primarily based upon a combination of two orthogonal approaches: DWT and NSCT. This combination can produce results comparable to that of the traditional NSCT without using too many scales of decomposition. As a result, the computational complexity incurred by the proposed framework is significantly less than that of the conventional NSCT. The basic steps involved in the proposed algorithm are:

1. Preprocessing.
2. Wavelet decomposition.
3. Decomposition, fusion and inverse transform of the approximation coefficients in the non-subsampled contourlet domain domain.
4. Fusion and inverse wavelet transform of combined approximation and merged detail coefficients.

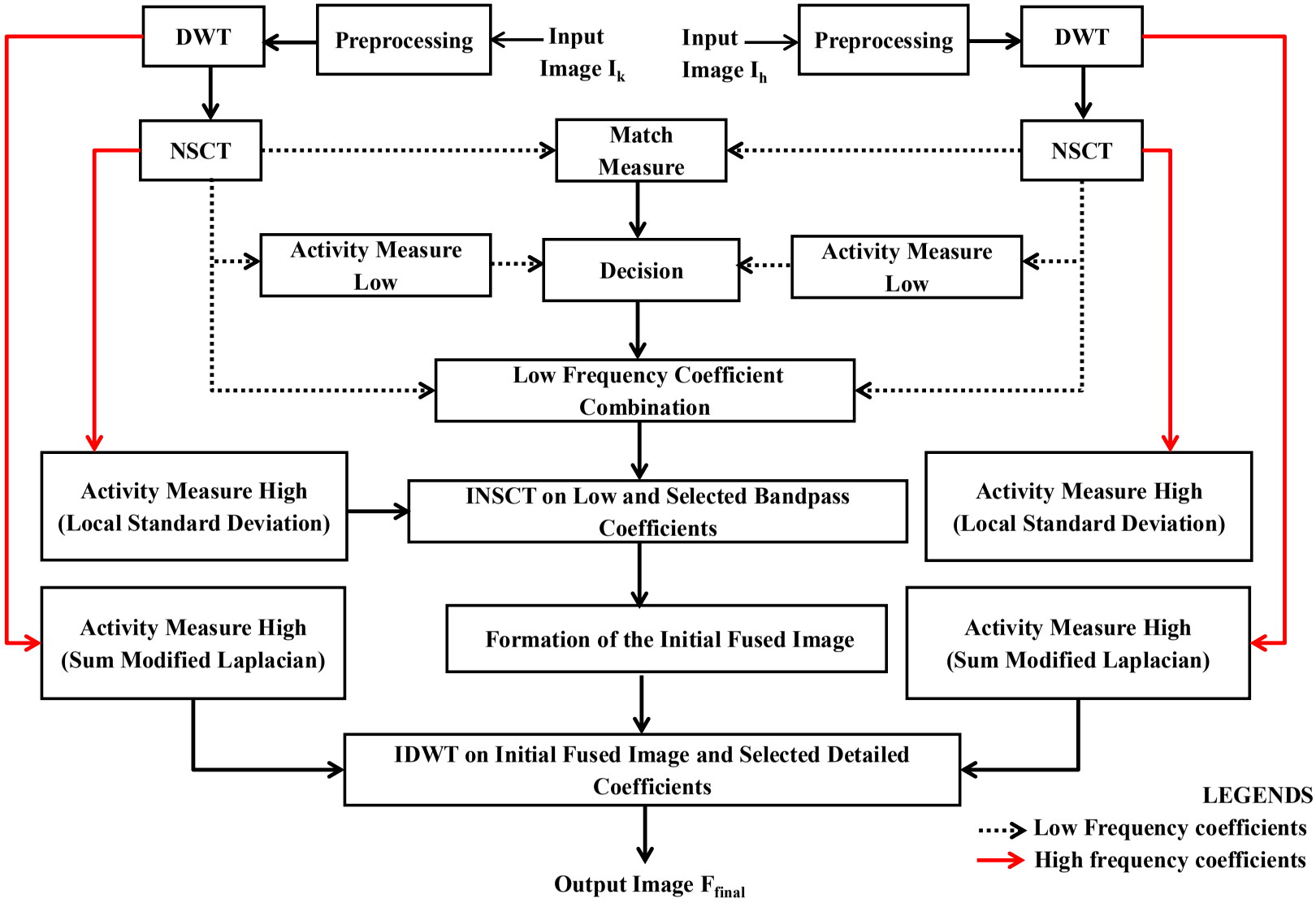

Figure 3.1: Proposed image fusion algorithm based on the architecture proposed by Piella [16].

3.2.1 Multi-scale decomposition and image fusion

Consider a pair of registered source images, say I_k and I_h. Initially, these source images are subjected to an auto-contrast enhancement mechanism [19] as part of the preprocessing stage. They are then subjected to a single scale wavelet decomposition by DWT. Here, the wavelet used is $daubechies - 8$ ($db8$). The resulting decomposition will result in approximation coefficients A_k, A_h and detail coefficients $D_k(i)$, $D_h(i)$, where $i = 1...3$, correspond to vertical, horizontal and diagonal frequency bands. The approximation coefficients are further decomposed into low frequency coefficients L_k, L_h and bandpass coefficients $B_k(g, m)$, $B_h(g, m)$, respectively, using NSCT. These coefficients are then combined using a fusion rule based on a match measure $M(L)$ and an adaptive threshold $T(L)$ to form the primary fused image $F_{initial}$. Finally, the consolidated output image F_{final} is obtained by taking the inverse wavelet transform of $F_{initial}$ and the merged detail coefficients $F(D_i)$. The basic steps involved in the proposed image fusion algorithm are described in the flowcharts (Fig. 3.2 and Fig. 3.3).

3.2.2 Formation of the initial fused image

The initial fused image $F_{initial}$ is obtained by reconstructing fused coefficients $F(L)$ and $F(B(g, m)$ in the NSCT domain.

$$F_{initial} = N^{-1}\{F(L), F(B(g, m))\} \tag{3.1}$$

Fusion of low frequency component

The low frequency coefficient fusion function $F(L)$ can be written as follows:

$$F(L) = \begin{cases} L_w & M(L) \geq T(L) \\ L_s & M(L) < T(L) \end{cases} \tag{3.2}$$

Here, $M(L)$ is a match measure [20], which indicates the degree of similarity between the two images at a spatial location (x, y) and is given by

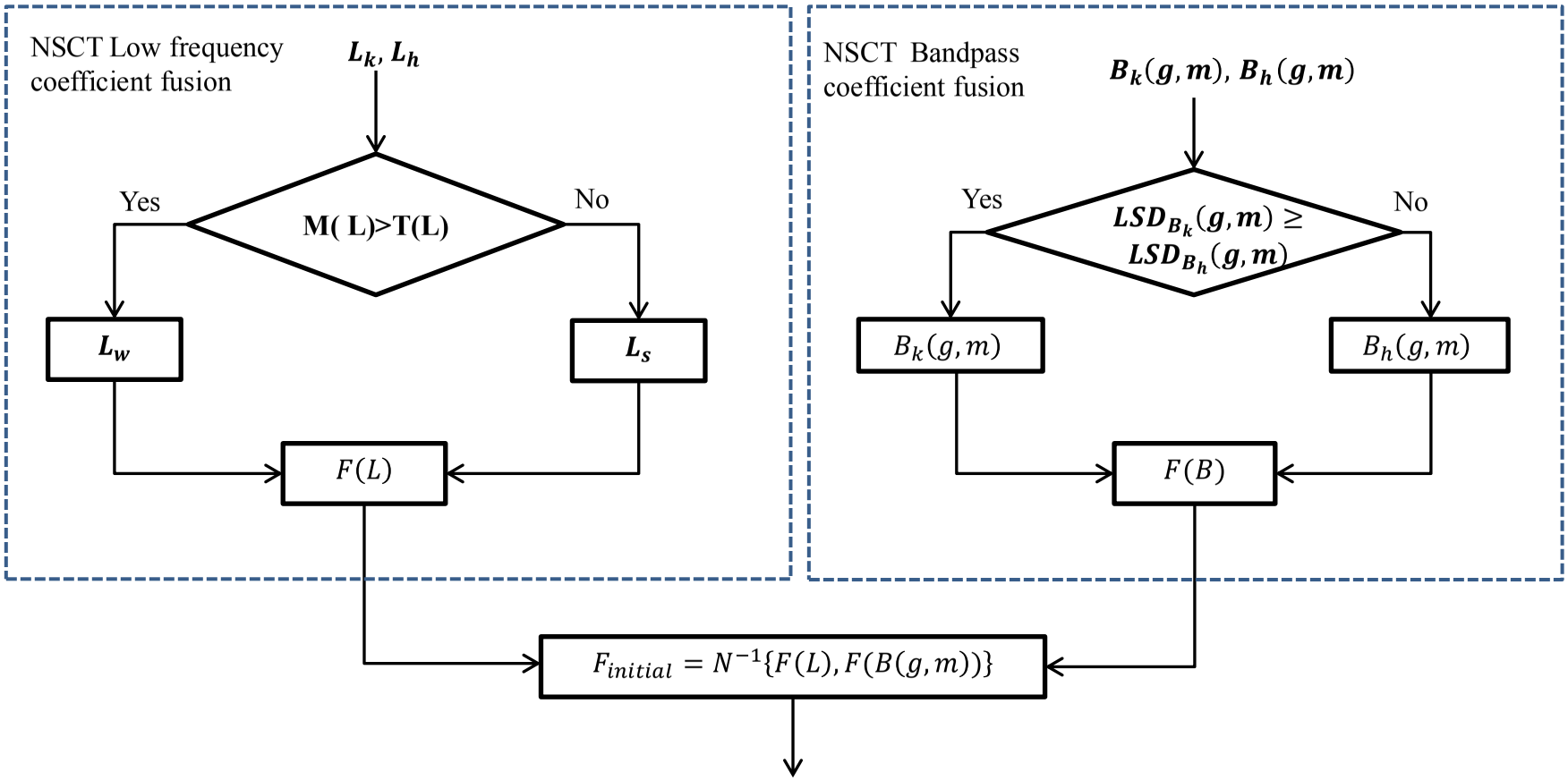

Figure 3.2: Formation of the initial fused image

.

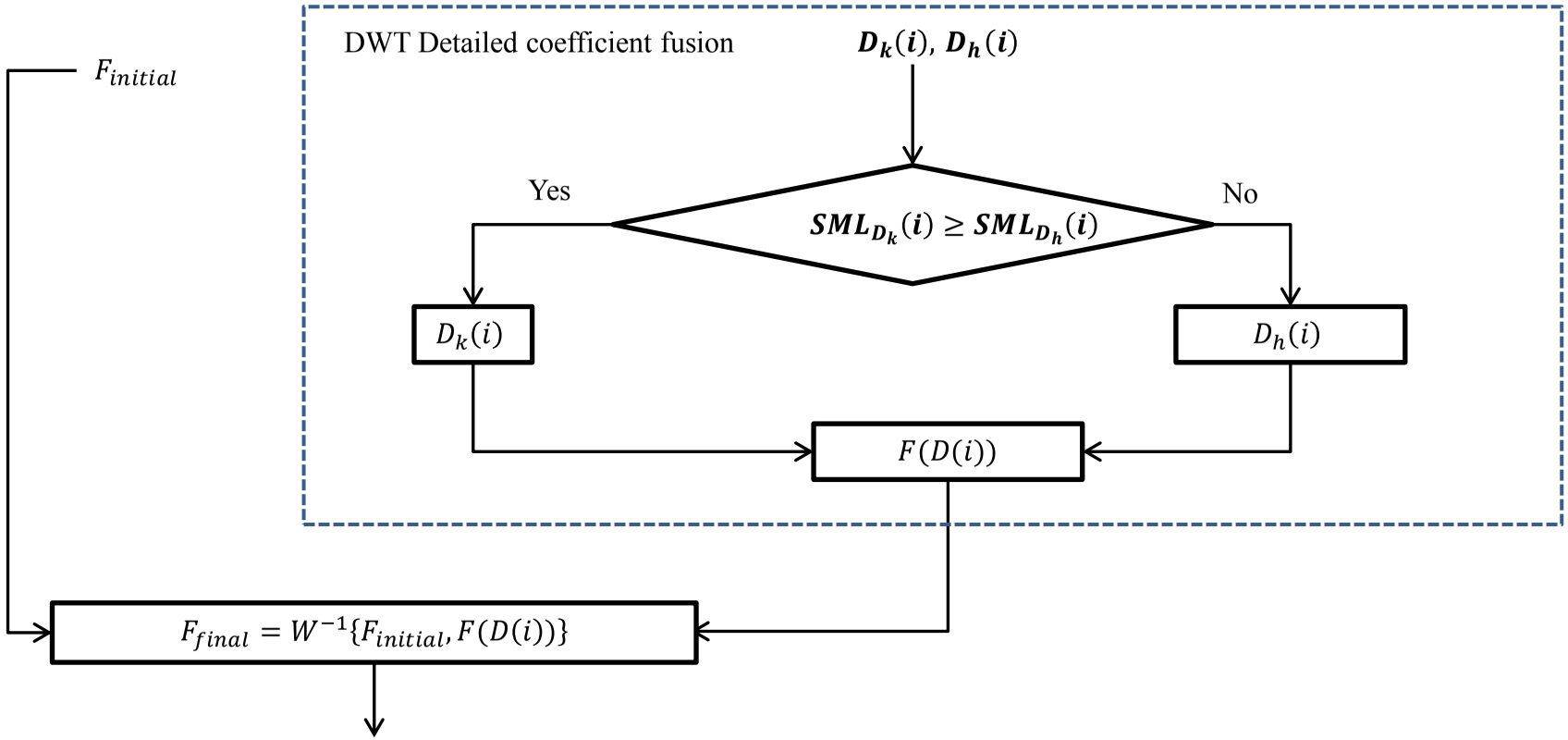

Figure 3.3: Formation of the final fused image

.

$$M(L) = \frac{2\sum_m \sum_n A_k(x+m, y+n) A_h(x+m, y+n)}{E(L_k) + E(L_h)} \tag{3.3}$$

where $E(L_c)$ is the local energy calculated within a window $w(m, n)$ and $c = k$ or h, and

$$E(L_c) = \sum_m \sum_n A_c(x+m, y+n)^2 w(m, n) \tag{3.4}$$

The size of $w(m, n)$ varies from $3 * 3$ for a $256 * 256$ source image to $5 * 5$ and $7 * 7$ for subsequent larger images. Also, local energies $E(L_k)$ and $E(L_h)$ are used to determine the adaptive threshold $T(L)$.

$$T(L) = \begin{cases} \frac{E(L_k)}{E(L_k)+E(L_h)}, & E_{L_k} \geq E_{L_h} \\ \frac{E(L_h)}{E(L_k)+E(L_h)}, & E_{L_k} < E_{L_h} \end{cases} \tag{3.5}$$

Depending upon the condition mentioned in (3.2), the low frequency coefficients L_w and L_s are either fused by weighted averaging L_w or selection L_s schemes:

$$L_w = \begin{cases} (A_k)(\Phi) + A_h(1-\Phi), & E_{L_k} \geq E_{L_h} \\ (A_k)(1-\Phi) + A_h(\Phi), & E_{L_k} < E_{L_h} \end{cases} \tag{3.6}$$

where ϕ is the weight smoothing function [20] given by

$$\Phi = \frac{1}{2} + \frac{1}{2}\left(\frac{1 - M(L)}{1 - T(L)}\right) \tag{3.7}$$

and

$$L_s = \begin{cases} (A_k), & E_{L_k} \geq E_{L_h} \\ (A_h), & E_{L_k} < E_{L_h} \end{cases} \tag{3.8}$$

The fused frequency component $F(L)$ obtained by the fusion rules mentioned in (3.6) and (3.8) ensure that a significant amount of the background information present in the low-frequency coefficients L_k and L_h are adequately preserved.

Fusion of the bandpass coefficients

The bandpass coefficients $B_k(g,m)$ and $B_h(g,m)$ contain features such as contours, edges and smooth regions of the input images. In addition, the human visual system is extremely sensitive to changes in contrast between the borders separating these features. Hence, a conventional activity measure based on the maximum absolute value will not be effective in detecting these variations Therefore, we use an activity measure, $LSD_{B_c}(g,m)$ (where $c = k$ or h), that can produce coefficients of large magnitude at the border regions by calculating the local standard deviation (LSD) [21]:

$$F(B) = \begin{cases} B_k(g,m), & LSD_{B_k}(g,m) \geq LSD_{B_h}(g,m) \\ B_h(g,m), & LSD_{B_k}(g,m) < LSD_{B_h}(g,m) \end{cases} \tag{3.9}$$

where $F(B) = F(B(g,m))$, g is the scale of decomposition and m is the number of directional subbands.

3.2.3 Formation of the final fused image

The final fused image F_{final} is obtained by reconstructing the initial fused image $F_{initial}$ and the detailed coefficients $F(D(i))$ in the wavelet domain:

$$F_{final} = W^{-1}\{F_{initial}, F(D(i))\} \tag{3.10}$$

Here, the sum-modified Laplacian (SML) [15] is used to fuse high quantity edge information present in the detailed sub bands $D_k(i)$ and $D_h(i)$. This is mainly due to its effectiveness in identifying notable features such as the edges and lines [22]. The mathematical expressions for the calculation of SML are shown below.

$$\begin{aligned} \nabla^2_{ML}\ f(x,y) = {} & |\,2f(x,y) - f(x-step,y) - f(x+step,y)| \\ & + |\,2f(x,y) - f(x,y-step) + f(x,y+step)| \end{aligned} \tag{3.11}$$

$$SML = \sum_{i=x-N}^{i=x+N} \sum_{j=y-N}^{j=y+N} \nabla^2_{ML} f(i,j) \tag{3.12}$$

where ∇^2_{ML} is the modified laplacian and *step* is the variable spacing between the pixels to calculate ∇^2_{ML}. Further details concerning the calculation of the variable step, which is initialized as unity in our proposed algorithm, can be found in [15].

$$F(D(i)) = \begin{cases} D_k(i), & SML_{D_k}(i) \geq SML_{D_h}(i) \\ D_h(i), & SML_{D_k}(i) < SML_{D_h}(i) \end{cases} \tag{3.13}$$

where $i = 1...3$, correspond to vertical, horizontal and diagonal frequency bands.

The above procedure is summarized in the form of the following steps.

Step 1: Wavelet decomposition of the input images:
$W(I_k) = [A_k, D_k]$ and $W(I_h) = [A_h, D_h]$

Step 2: NSCT decomposition of the approximation coefficients:
$N(A_k) = [L_k, B_k]$, $N(A_h) = [L_h, B_h]$

Step 3: Formation of the initial fused image $F_{initial}$.

Step 4: Formation of the final fused image F_{final}.

3.3 Experimental results and analysis

The performance of the proposed algorithm is now compared against the fusion results obtained by applying Laplacian [7], HOSVD [2], and the conventional NSCT [11] methods. In the case of NSCT in our presented framework, *pyrexc* and *vk* [23] are chosen as the *pyramidal* and *directional* filter banks, respectively, and the decomposition scales are configured based upon the resolution depth and the type of the source images. The optimum decomposition scale g and the number of directive bandpass subbands m of the NSCT used in our presented framework are shown in Table 3.1.

In the case of the Laplacian approach, the low and high-frequency coefficients are fused by the absolute maximum and averaging schemes. In the conventional NSCT method, the scale configurations are initialized as shown in Table 3.1. Remaining settings for the Laplacian and the traditional NSCT approaches are kept the same as that of the source codes available in [24, 25]. The default settings presented in [26] are followed for the

Table 3.1: Non-subsampled contourlet transform scale configurations

Input Images and Size	(A1, A2) 256*256	(B1, B2) 256*256	(C1, C2) 512*512	(D1, D2) 1280*960
g	1	2	3	3
m	16	2,4	2,4,8	2,4,16

HOSVD approach. For quantitative and qualitative evaluation of the proposed algorithm, simulations are performed on four pairs of images.

For the first pair of images, namely, multi-modal medical images Fig. 3.4 (A1, A2), it is seen from Fig. 3.4 that the fused images B1, B2, and C1 suffer from numerous issues such as loss of soft-tissue content, reduced contrast and unclear edges of the brain boundaries. The fused image C2 obtained by our proposed method looks more visually pleasing and has clear distinguishable edges, especially at the brain boundaries. In the case of multi-modal remote-sensing images Fig. 3.5 (A1, A2), the fused image C2 obtained by using the proposed algorithm has high texture content and perfectly visible edges. The fused image B1 obtained by the Laplacian method suffers from the lack of texture content, whereas B2 and C1 contain unclear edges. In cases of multi-focal clock Fig. 3.6 (A1, A2) and book images Fig. 3.7 (A1, A2), fused images C2 obtained using the proposed image fusion algorithm are highly focused. Moreover, they also have bright appearances due to the enhanced contrast.

For quantitative analysis of the results, the most commonly used image fusion performance measures, namely, the $Q_{AB/F}$ metric [17] and entropy are employed. In addition, these performance measures do not require a reference image or an original image. These quantitative performance measures should be viewed in a combination rather than individually to get an exact idea on the effectiveness of the various fusion algorithms. The quantitative results obtained from the simulations are shown in Table 3.2. From this table, it can be observed that our proposed method has the best $Q_{AB/F}$ values for the book and clock images and has the second best $Q_{AB/F}$ values for the medical and remote-sensing images. Furthermore, the proposed method has the highest entropy content than its counterparts, thereby indicating the effectiveness of our algorithm in preserving the information content.

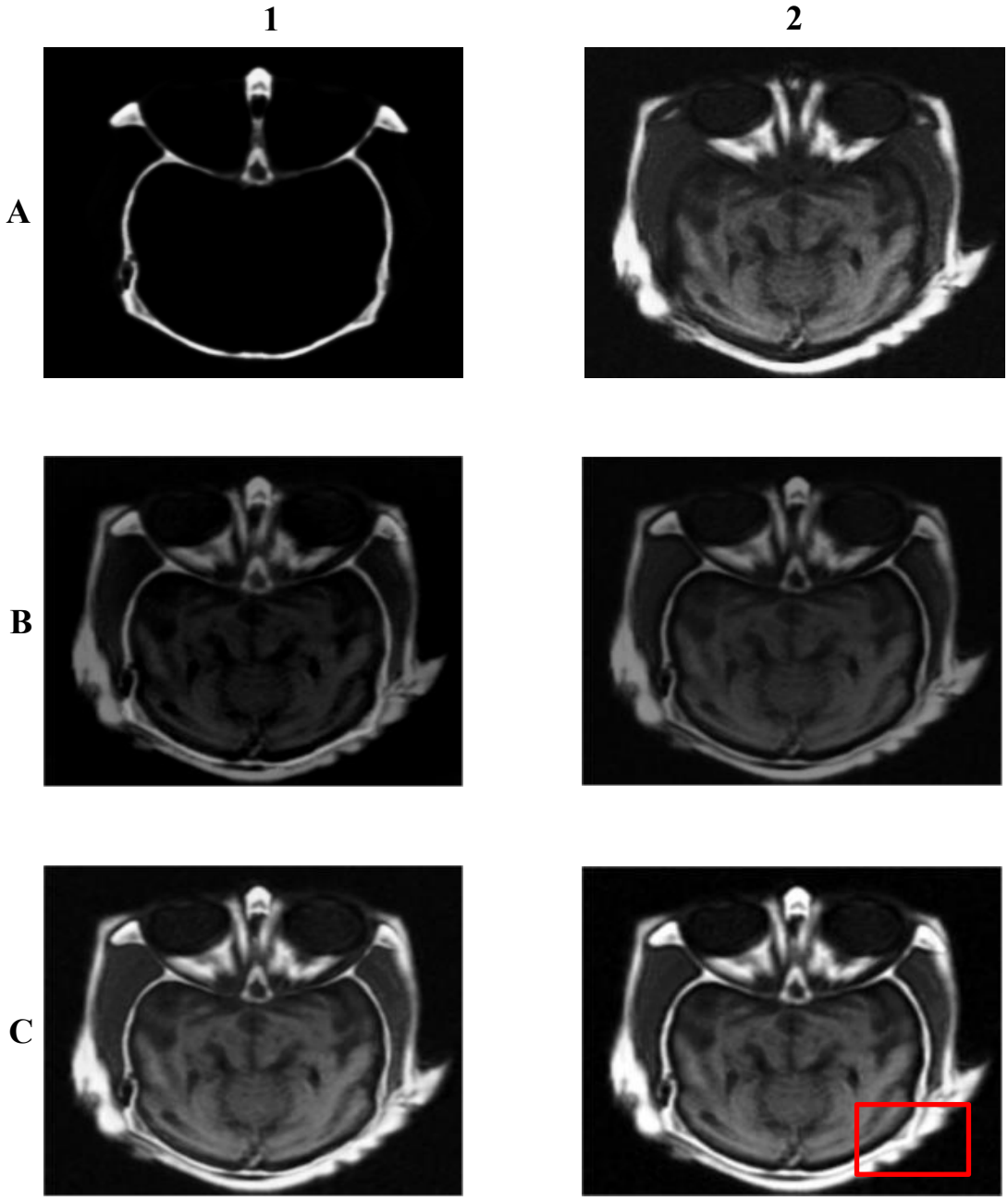

Figure 3.4: Qualitative analysis: Source image [3]: Medical images (A1 - CT, A2 - MRI), Fused Results: [7] Laplacian (B1), [2] HOSVD fused (B2), [11] NSCT (C1) , Proposed method (C2)

.

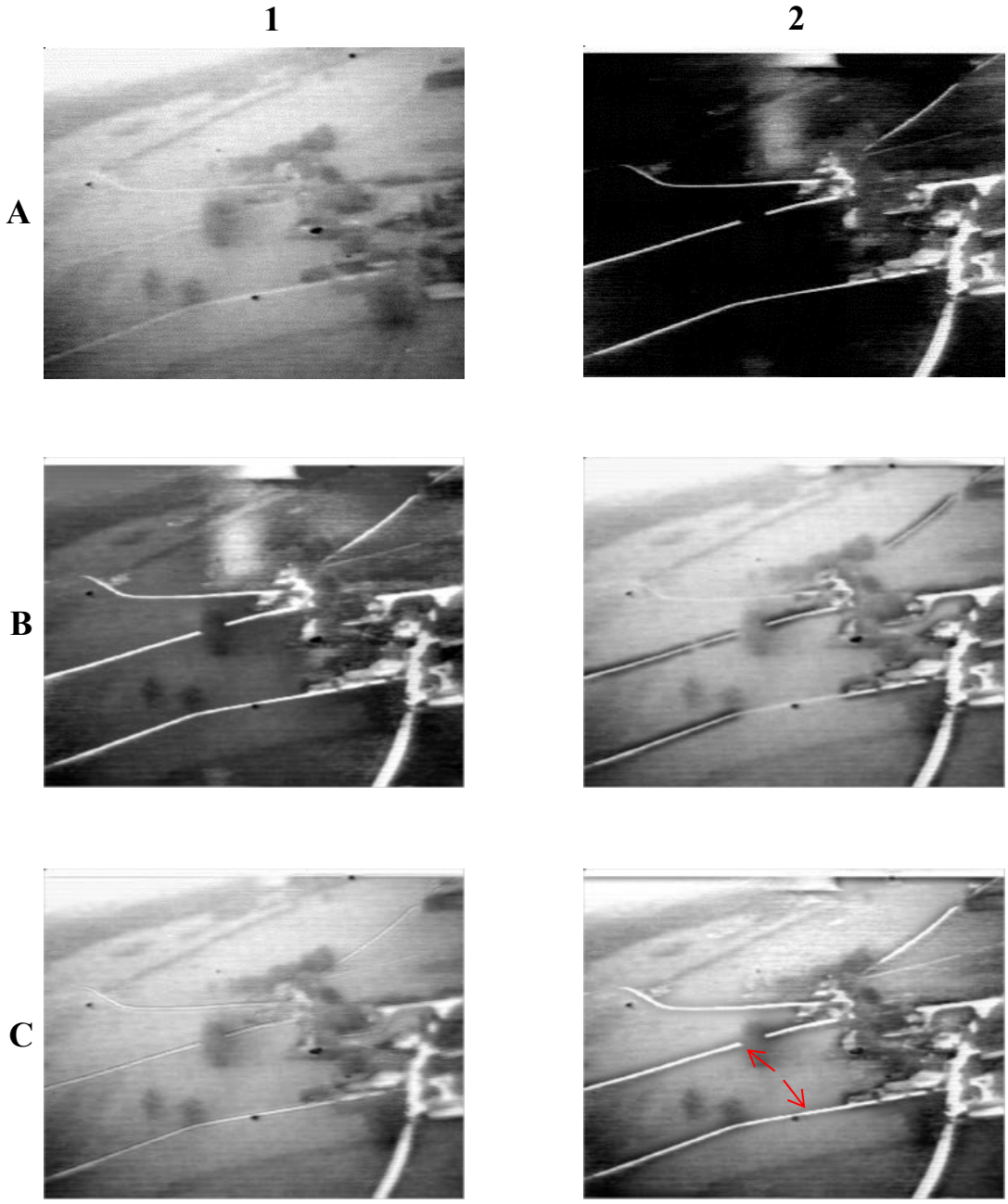

Figure 3.5: Qualitative analysis: Source images [3]: Remote sensing (A1 - LLTV image, A2 - FLIR image), Fused Results: [7] Laplacian (B1), [2] HOSVD fused (B2), [11] NSCT (C1) , Proposed method (C2)
.

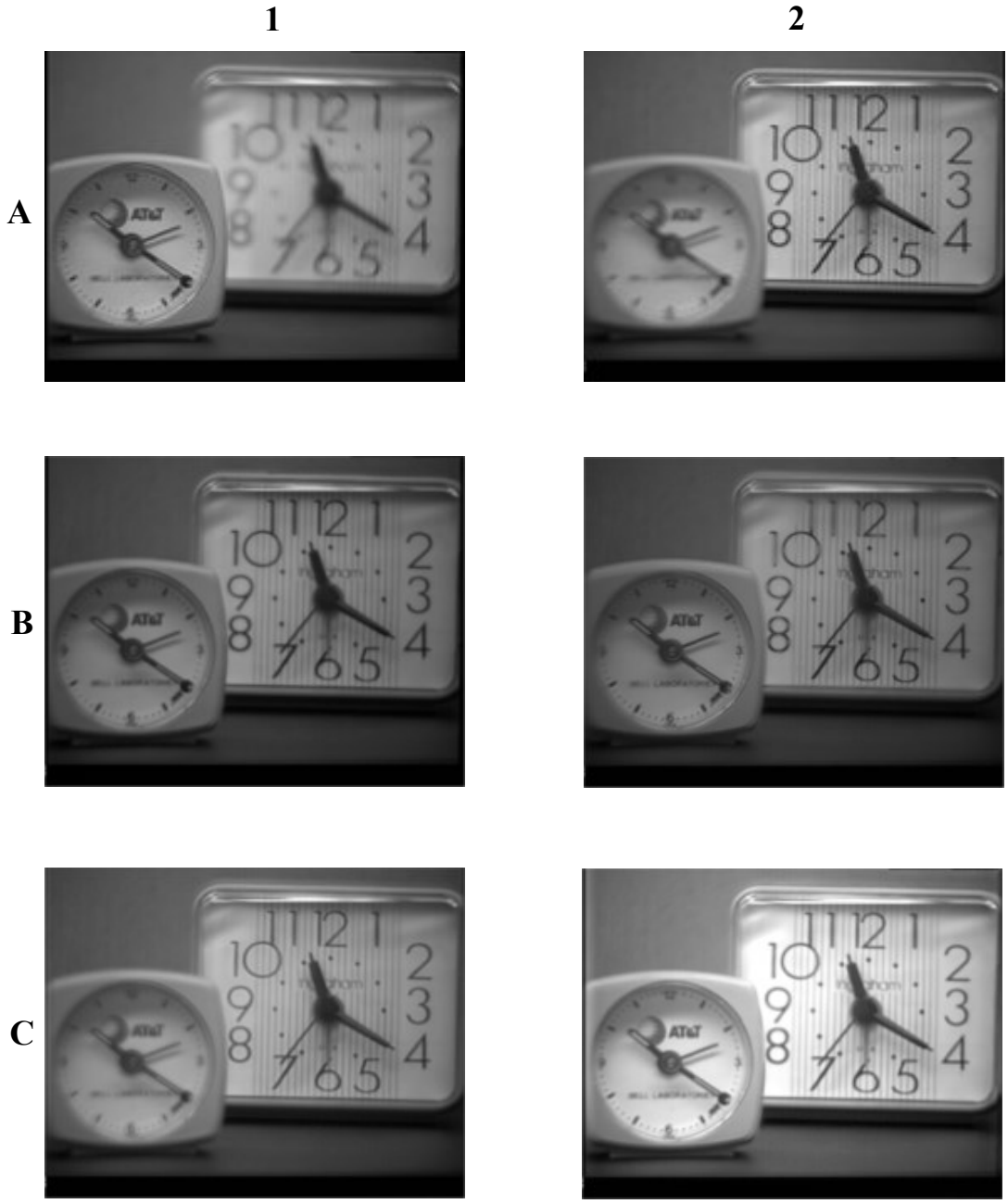

Figure 3.6: Qualitative analysis: Source images Multi-Focus Images (A1 - left focused, A2 - right focused), Fused Results: [7] Laplacian (B1), [2] HOSVD fused (B2), [11] NSCT (C1) , Proposed method (C2)
.

Figure 3.7: Qualitative analysis: Source images Multi-Focus Images (A1 - left focused, A2 - right focused), Fused Results: [7] Laplacian (B1), [2] HOSVD fused (B2), [11] NSCT (C1) , Proposed method (C2)
. only if needed

Table 3.2: Quantitative performance of the proposed image fusion algorithm

Input Images	**$Q_{AB/F}$**				**Entropy**				**Execution Time (T) in Seconds**			
	[7]	**[2]**	**[11]**	**Proposed Method**	**[7]**	**[2]**	**[11]**	**Proposed Method**	**[7]**	**[2]**	**[11]**	**Proposed Method**
Medical	0.7320	**0.7868**	0.7569	0.7821	5.2931	5.8384	5.9723	**6.4673**	0.0220	67.22	0.808	2.1479
Remote sensing	**0.6046**	0.5051	0.4032	0.5418	7.2683	7.6389	7.5404	**7.7499**	0.0226	65.55	6.914	2.0324
Clock	0.6620	0.6367	0.6530	**0.6725**	7.0652	6.9760	7.0060	**7.7479**	0.0790	275.35	85.520	5.1060
Book	0.6858	0.7096	0.6148	**0.7112**	7.3109	7.2807	7.3437	**7.3974**	0.3723	1246.7	732.391	22.109

Platform used: Matlab version R2012 a, **System specifications:** Intel Pentium P6200, 6GB DDR3 Memory.
Methods [7], [2] and [11] corresponds to Laplacian, HOSVD and conventional NSCT, respectively

In terms of the execution time, for larger images the proposed method remains much faster than the conventional NSCT and the HOSVD approaches. This is mainly due to the initial decomposition by DWT and the computationally- efficient fusion rules. However, the execution time of the proposed algorithm is slightly more than that of the Laplacian method. This is understandable, as our present method is based upon a combination of multi-scale transforms. The complexity of our algorithm remains intact even with changes in scales of decomposition unlike in the case of the traditional NSCT. In the traditional NSCT, the execution time is directly dependent on its scale of decomposition. This is quite evident as the execution time of the conventional NSCT for the fused image B5 is much greater than that of the fused image A5, even though both the image pairs are of the same size.

3.4 Summary

In this chapter, a novel image fusion algorithm capable of producing high-quality fused images even with a two-scale decomposition has been proposed. It has been shown that, in general, our approach is highly successful in preserving features such as tissue content, texture and edges from various multi-modal sensors. Furthermore, our algorithm has been shown to be capable of producing high quality fused images with lesser computation time. Furthermore, the complexity of our algorithm remains intact even for images of large dimensions unlike the existing image fusion methods. The aforementioned merits qualify the proposed image fusion technique as an ideal candidate for an application like real-time multi-modal surveillance using video fusion.

Chapter 4

Camouflaged target detection using real-time video fusion algorithm based on multi-scale transforms

In this chapter, we propose a novel video fusion algorithm for real-time detection of camouflaged targets. Initially, the targets are detected by using a novel target detection method by applying conventional image thresholding methods in the wavelet domain. The thermal information corresponding to the detected targets are transferred to the output fused image along with the background information by using novel region-based fusion rules. The dominance of the proposed algorithm in terms of the detection of the camouflaged targets is visibly evident from the quantitative and qualitative results of the output fused video.

4.1 Introduction

Multi-modal imaging sensors are an integral part of any surveillance system. These sensors provide a wide variety of data such as infrared and visible background information

contained in a scene. The inputs from the infrared sensors can be used to recognize and keep track of various targets such as humans, moving objects and animals, even from a highly crowded background. This information can be seen only in the infrared spectrum and cannot be perceived in the visible spectrum. This infrared data when combined together with the visible information provides us with an output having both thermal as well as visible geometrical features, thereby improving the various essential capabilities of a surveillance system such as object detection and tracking.

In general, infrared images provide a clear and distinct view of various objects present in a dark environment. On the down side, these images are very challenging to understand as all the background information contained in the scene is totally lost. In the case of visible images, all the background information will be adequately retained, but thermal data such as heat signatures will be completely lost. One typical application that requires a combination of both infrared and thermal imaging sensors are camouflaged target detection.

Camouflage is a term used to describe a situation in which a visible entity is intentionally concealed from the environment through color toning of the visible object and the background. The foremost challenge involved behind camouflage target detection is to disintegrate the hidden target as well as to retain the various geometrical structures and normal colors contained in the background. The camouflaged target in the visible spectrum can be easily localized by exploiting the corresponding thermal information present in the infrared video. The most commonly employed solution for the above-mentioned problem is the fusion of the infrared and visible videos. The most commonly observed issues that arise after the fusion of these videos is the low contrast of the output video with the conventional fusion methods [7], [2], [11]. This is mainly caused due to the extremely low intensity values of the corresponding pixels present in the infrared video.

To address the above-specified issue, we extend in this chapter the novel framework described in Chapter 3. The main contribution of the proposed algorithm is the formulation of the unique region-based fusion rules. In addition, an effective target detection and highlighting mechanism is developed by applying conventional image thresholding techniques

in the discrete wavelet domain.

4.2 Proposed region-based video fusion algorithm

The architecture of the proposed region-based fusion algorithm is as shown in Fig. 4.1. The core idea behind the proposed algorithm is the usage of novel region-based fusion rules along with an improved coefficient decomposition framework based on the multi-scale transforms to obtain a real-time fused output.

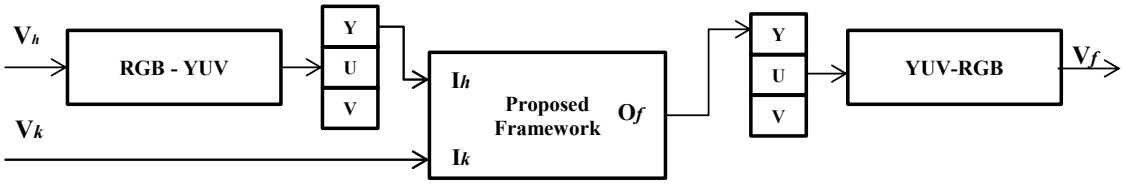

Figure 4.1: Proposed image fusion algorithm based on the architecture proposed by Piella [16].

4.2.1 RGB to YUV Conversion

Initially, the visible video input V_h is converted from the RGB color space to the YUV color space. This is mainly due to variances between the characteristic vision perception and the distance amid the two points present in the RGB color space. Due to this, the various attributes like hue saturation and brightness cannot be obtained from the RGB data. In the case of YUV color space, Y component corresponds to the luma sensitivity, U and V corresponds to the chroma perception. As this color space comprises of nonlinear luma/chroma, each constituent is free from each other. RGB to YUV conversion [27] of the input visible video V_h as follows:

Intially, we assume the following constants

$$\begin{aligned} K_R = 0.299, K_B = 0.114, K_G = 1 - K_R - K_B = 0.587 \\ U_{max} = 0.436, V_{max} = 0.615 \end{aligned} \tag{4.1}$$

where K_R, K_G and K_B are the weighted values of the R, G, and B components. These components are added to obtain the Y component, which is the overall measure of luminance. The other two components U and V are calculated by taking scaled differences of the B and R components from the Y component as given below.

$$\begin{aligned} Y &= K_R R + K_G G + K_B B \\ U &= U_{max}\frac{B-Y}{1-K_B} \approx 0.492(B-Y) \\ V &= V_{max}\frac{R-Y}{1-K_R} \approx 0.877(R-Y) \end{aligned} \tag{4.2}$$

Substituting (4.1) in (4.2), we obtain the final expression as

$$\begin{bmatrix} Y_{vis} \\ U_{vis} \\ V_{vis} \end{bmatrix} = \begin{bmatrix} 0.2999 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R_{vis} \\ G_{vis} \\ B_{vis} \end{bmatrix} \tag{4.3}$$

where $Y_{vis} = Y$, $U_{vis} = U$, $V_{vis} = V$, $R_{vis} = R$, $G_{vis} = G$ and $B_{vis} = B$.

4.2.2 Multi-scale decomposition and video fusion

Consider a pair of registered input thermal and visible videos as V_k, V_h respectively. Initially, the visible video V_h undergoes the above mentioned RGB to YUV conversion. After the conversion, the thermal video V_k and the Y component of the visible video V_h are correspondingly assumed as I_k and I_h. They are then subjected to a single scale wavelet decomposition by DWT. Here, the wavelet used is $daubechies - 8$ ($db8$). The resulting decomposition will result in approximation coefficients A_k, A_h and detail coefficients $D_k(i)$, $D_h(i)$, where $i = 1...3$, correspond to vertical, horizontal and diagonal frequency bands. The approximation coefficients are further decomposed into low frequency coefficients L_k, L_h and bandpass coefficients $B_k(g,m)$, $B_h(g,m)$, respectively using NSCT. These coefficients are then combined using a fusion rule based on the conventional thresholding techniques of Kapur et al. [14] and Otsu [13] to form the primary fused image frame $F_{initial}$. Finally, the consolidated output image F_{final} is obtained by taking the inverse wavelet transform of $F_{initial}$ and the merged detail coefficients $F(D_i)$.

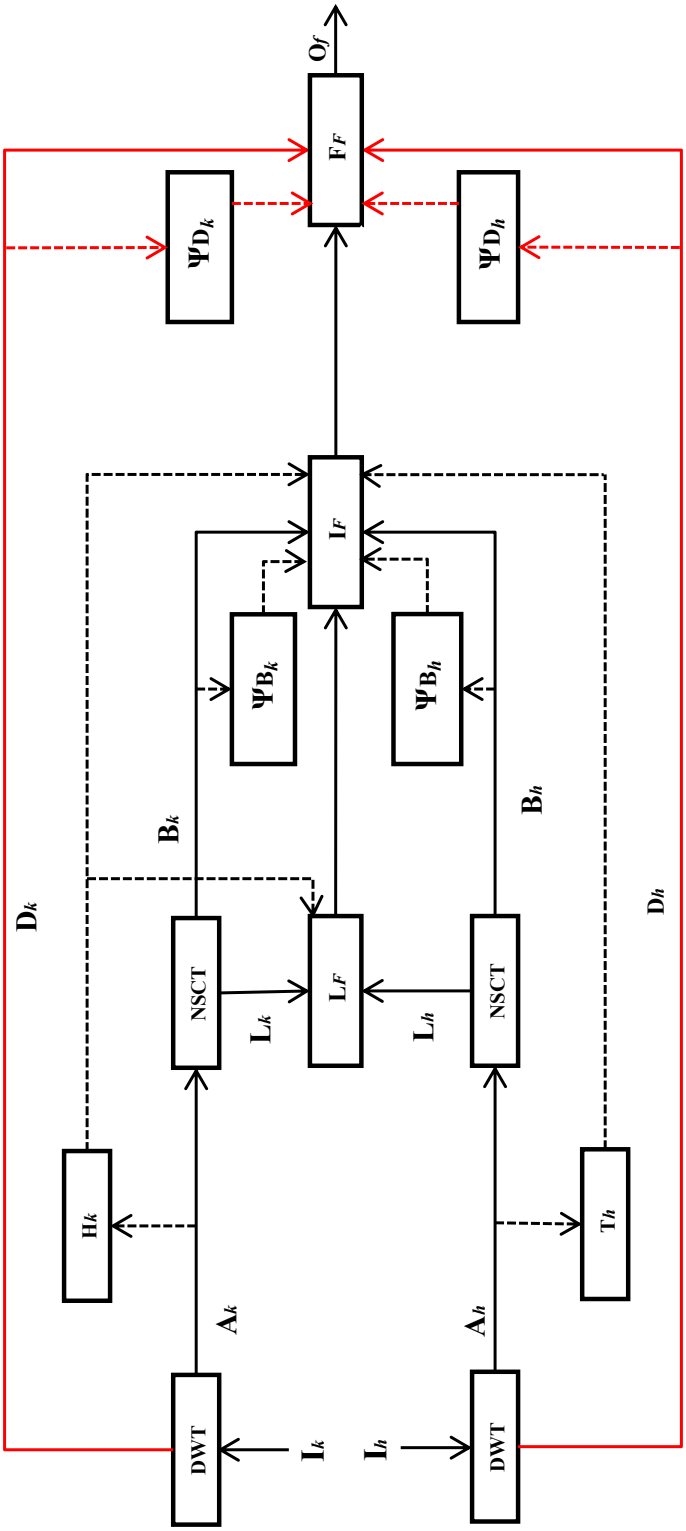

Figure 4.2: Proposed region-based image fusion architecture.

4.2.3 Formation of the initial fused frame

The low frequency NSCT coefficients L_k and L_h are suitably fused using a novel target detection method and simple averaging techniques. The unique camouflage target detection technique is based on the application of the conventional thresholding methods of Kapur et al. [14] and Otsu [13]. These thresholding techniques are employed on the low frequency discrete wavelet coefficient of the infrared video coefficients A_k and A_h to detect the target as well as active regions involved in the particular video frame to form the initial fused frame $F_{initial}$.

Fusion of the low frequency component

In our proposed video fusion framework, we formulate a relatively simple region-based fusion rule for combining the low frequency NSCT components L_k and L_h. The core mechanism behind the proposed fusion rule is to apply the threshold H_k of Kapur et al. [14] in the transform domain rather than the conventional approaches based in the spatial domain. The details and various steps involved in the calculation of the threshold H_k is discussed below.

Threshold H_k:

The threshold of Kapur et al. [14] is primarily based on the entropies of the input image. In this method, the foreground and the background of an input image are considered as two different sources. The optimum threshold condition of this method occurs, when the sum of the foreground and background classes is at the maximum value. The foreground and background entropies are calculated as follows:

$$H_{fg}(T) = \sum_{k=0}^{T} \frac{p(k)}{P(T)} log \frac{p(k)}{P(T)} \tag{4.4}$$

$$H_{bg}(T) = - \sum_{g=T+1}^{G} \frac{p(g)}{P(T)} log \frac{p(g)}{P(T)} \tag{4.5}$$

$$H_k = argmax \left[H_{fg}(T) + H_{bg}(T) \right] \tag{4.6}$$

By applying the threshold H_k using (4.6) on the approximate coefficient A_k of the input thermal video V_k, the regions containing camouflaged targets are identified. The corresponding active regions in the NSCT low frequency thermal component L_k are then combined with NSCT low frequency visible component L_h to obtain the low frequency NSCT fused component F_L. This combination is guided by the active regions detected by the conventional Kapur threshold H_k from the approximate thermal coefficient A_k.

$$F_L = (A_k)(\Phi_R) + A_h(1 - \Phi_R) \tag{4.7}$$

where Φ_R is the active regions detected by the Kapur threshold H_k from the low frequency thermal component L_k.

Fusion of the bandpass coefficients

The bandpass coefficients $B_k(g, m)$ and $B_h(g, m)$ are rich in significant geometrical features such as contours, edges and smooth regions of the thermal and visible videos. Here, we propose a novel fusion rule based on the conventional Otsu thresholding [13] and Sum-Modified Laplacian (SML) [15] for the fusion of these bandpass coefficients. Initially, Otsu thresholding is applied to the approximate wavelet component of the visible video A_k to obtain the various active regions. A detailed overview of SML can be found in the chapter 3 and the various steps involved in the calculation of Otsu threshold is discussed below.

Threshold T_h:
The main principle involved behold Otsu thresholding [13] is to explore for a threshold which shrinks the variance among the classes to the minimum. The intra-class variance can be further defined as the sum of the variances of the two classes as shown below:

$$\sigma_w^2(t) = \phi_1(t)\sigma_1^2(t) + \phi_2(t)\sigma_2^2(t) \tag{4.8}$$

where w_i, t and σ_i^2 corresponds to the probabilities, threshold and variances of the two classes respectively. As per the basic notion of Otsu thresholding, minimizing the variance among the intra-class is as same as that of maximizing the variance among the inter-class.

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = w_1(t)\, w_2(t)\, [\mu_1(t) - \mu_2(t)]^2 \tag{4.9}$$

where w_I and μ_i corresponds to the probabilities and the means of the two classes respectively. The required class probability $w_1(t)$ is then obtained as shown below from the histogram of the input image as t.

$$w_i(t) = \sum_{0}^{t} p(i) \tag{4.10}$$

where $i = 1, 2$ and the mean of the class $\mu_1(t)$ is calculated as follows:

$$\mu_1(t) = \left[\frac{\sum_0^t p(i)x(i)}{w_1} \right] \tag{4.11}$$

where x_i corresponds to the i^{th} value of the histogram bin.
Correspondingly, $w_2(t)$ and $\mu_2(t)$ are estimated by taking values greater than t on the right hand side of the histogram. The various steps are involved as follows.

1. Estimate the histogram and various probabilities corresponding to each intensity level.
2. Initialize the values of $\omega_i(0)$ and $\mu_i(0)$.
3. Allocate the various possible threshold values from $t = 1 \ldots$ maximum intensity.
   (a) Update ω_i and μ_i.
   (b) Compute $\sigma_b^2(t)$.
4. Desired threshold corresponds to the maximum $\sigma_b^2(t)$.
5. Compute the two maxima (and two corresponding thresholds). $\sigma_{b1}^2(t)$ is the greater max value and $\sigma_{b2}^2(t)$ is greater or equal maximum value.
6. Preferred threshold = $\frac{threshold_1 + threshold_2}{2}$.

Furthermore, a negation is performed on the decision map $T_h(i, j)$ to obtain the active regions containing the camouflaged target as shown below.

$$\phi_B(i, j) = 1 - T_h(i, j) \tag{4.12}$$

In addition to the decision map $\phi_B(i,j)$, an another activity measure $\Psi_S(i,j)$ is also calculated using

$$\Psi_S(i,j) = \begin{cases} 1, & SML_{D_k}(i) \geq SML_{D_h}(i) \\ 0, & SML_{D_k}(i) < SML_{D_h}(i) \end{cases} \tag{4.13}$$

where $S = k$ or h.

These two decision maps $\phi_B(i,j)$ and $\Psi_S(i,j)$ ensure that all the active regions of the thermal video are preserved and transferred along with the visible background to the fused bandpass output $F(B)$. The fusion of bandpass coefficients $B_k(g,m)$ and $B_h(g,m)$ can be carried out by any of the two following modes.

1. **Normal mode**

   In this mode, the band pass coefficients are fused using (4.12) and (4.13) as shown below.

$$F(B) = \begin{cases} B_k(g,m), & \phi_B(i,j) == 1 \ \& \ \Psi_S(i,j) == 1 \\ B_h(g,m), & \text{for all other cases} \end{cases} \tag{4.14}$$

2. **Target highlighting mode**

   In this mode, the camouflaged target target can be artificially highlighted by combining the Kapur threshold H_k along with the decision maps from (4.12) and (4.13) as shown below.

$$F(B) = \begin{cases} (\alpha)\,(B_k(g,m)) + (H_k)\,(B_k(g,m)), & \phi_B(i,j) == 1 \ \& \ \Psi_S(i,j) == 1 \\ B_h(g,m), & \text{for all other cases} \end{cases} \tag{4.15}$$

   where $F(B) = F(B(g,m))$, g is the scale of decomposition, m is the number of directional subbands and α is an arbitrary constant kept at a value of 1.25.

4.2.4 Formation of the final fused frame

The final fused frame F_{final} is obtained by reconstructing the initial fused image $F_{initial}$ and the detailed coefficients $F(D(i))$ in the wavelet domain:

$$F_{final} = W^{-1}\{F_{initial}, F(D(i))\} \tag{4.16}$$

Here, the sum-modified Laplacian (SML) [15] is combined together with a parameter based on the ratio of local standard deviations (LSD) to fuse the high quantity edge information present in the detailed sub bands $D_k(i)$ and $D_h(i)$. The mathematical expressions for the fusion of detailed coefficients $F(D(i))$ are

$$F(D(i)) = \begin{cases} D_k(i), & (\phi_{L_k(i)})(SML_{D_k}(i)) \geq (\phi_{L_h(i)})(SML_{D_h}(i)) \\ D_h(i), & (\phi_{L_h(i)})(SML_{D_k}(i)) < (\phi_{L_h(i)})(SML_{D_h}(i)) \end{cases} \tag{4.17}$$

where $i = 1...3$, correspond to vertical, horizontal and diagonal frequency bands, and $\phi_{L_k(i)}$ and $\phi_{L_h(i)}$ are adaptive parameters based on the ratio of their local standard deviations $LSD_{B_k}(g,m)$ and $LSD_{B_h}(g,m)$. These parameters are calculated as shown below.

$$\phi_{L_k(i)} = \left\{ \frac{LSD_{B_k}(g,m)}{LSD_{B_k}(g,m)) + LSD_{B_h}(g,m)} \right. \tag{4.18}$$

and

$$\phi_{L_h(i)} = \left\{ \frac{LSD_{B_h}(g,m)}{LSD_{B_k}(g,m)) + LSD_{B_h}(g,m)} \right. \tag{4.19}$$

4.2.5 YUV to RGB Conversion

The final fused image frame F_{final} is converted back to RGB video [27] by combining it with its corresponding luma U_{vis} and chroma V_{vis} channels as shown below.
Inverting (4.2) to obtain the expression in terms of R, G and B of the fused output.

$$\begin{aligned} R &= F_{final} + V_{vis}\frac{1-K_R}{V_{max}} = F_{final} + \frac{V_{vis}}{0.877} \\ G &= F_{final} - U_{vis}\frac{K_B(1-K_B)}{U_{max}K_G} - V_{vis}\frac{K_R(1-K_R)}{V_{max}K_G} = F_{final} - 0.395U_{vis} - 0.581V_{vis} \\ B &= F_{final} + U_{vis}\frac{1-K_B}{U_{max}} = F_{final} + \frac{U_{vis}}{0.492} \end{aligned} \tag{4.20}$$

Substituting the constant values from (4.1) in (4.20), we obtain the final expression as

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0.000 & 1.1400 \\ 1 & -0.369 & 0.581 \\ 1 & 2.029 & 0.000 \end{bmatrix} \begin{bmatrix} F_{final} \\ U_{vis} \\ V_{vis} \end{bmatrix} \tag{4.21}$$

The above procedure is summarized in the form of the following steps.

Step 1: RGB to YUV conversion of the visible video V_h.

Step 2: Wavelet decomposition of the input frames (Y channel of visible and the thermal video):
$W(I_k) = [A_k, D_k]$ and $W(I_h) = [A_h, D_h]$

Step 3: NSCT decomposition of the approximation coefficients:
$N(A_k) = [L_k, B_k]$, $N(A_h) = [L_h, B_h]$

Step 4: Formation of the initial fused image $F_{initial}$.

Step 5: Formation of the final fused image F_{final}.

Step 6: Reconstructing YUV frames to obtain the fused RGB video.

4.3 Experimental results and analysis

The performance of the proposed video fusion algorithm is now compared against the fusion results obtained by applying Laplacian [7], HOSVD [2], and the conventional NSCT [11] methods. In the case of NSCT in our presented framework, *pyrexc* and *vk* [23] are chosen as the *pyramidal* and *directional* filter banks, respectively, and the decomposition scales are configured based upon the resolution depth and the type of source images. In our proposed framework, the optimum decomposition scale g is 4 and the number of directive bandpass subbands m present in each scale is 2,4,8 and 16.

In the case of the Laplacian approach, the low and high-frequency coefficients are fused by the absolute maximum and averaging schemes. In the conventional NSCT method, the scale configurations are initialized with 4 scales and directional bandpass subbands m in each scale as 2,4,8 and 16. Remaining settings for the Laplacian and the traditional NSCT approaches are kept the same as that of the source codes available in [24, 25]. The default settings presented in [26] are followed for the HOSVD approach. For quantitative and qualitative evaluation of the proposed algorithm, simulations are performed on four pairs

of randomly chosen thermal and visible image frames obtained from [3]. The dimensions of these input videos are kept at 512*512.

For the first pair of image frames, namely, 25^{th} image frames of thermal and visible videos Fig. 4.3 (A1, A2), it is seen from Fig. 4.3 that the fused images B1, B2, and C1 suffer from numerous issues such as loss of thermal content of the camouflaged target and unclear visible background . The fused image C2 obtained by our proposed method looks more visually pleasing and has a clear distinguishable background, especially at the boundaries of the target and the background. In the case of the 50^{th} image frames of thermal and visible videos Fig. 4.4 (A1, A2), the fused image C2 obtained by using the proposed algorithm has high thermal content and visible background features are adequately preserved. The fused image B1 obtained by the Laplacian method suffers from the loss of thermal content, whereas B2 and C1 contain unclear visible background. In cases of the 96^{th} image frames of thermal and visible videos Fig. 4.5 (A1, A2) and 100^{th} image frames of thermal and visible videos Fig. 4.6 (A1, A2), fused images C2 obtained using the proposed video fusion algorithm successfully preserves the visible background information as well as adequately expose the hidden target. In addition, the fused output obtained by the proposed method in the target highlighting mode (Fig. 4.7) clearly distinguishes the hidden target from the background.

For quantitative analysis of the results, the $Q_{AB/F}$ metric [17] and entropy are employed. Details concerning these measures can be found in Chapter 2. The quantitative results obtained from the simulations are shown in Table 4.1. From this table, it can be observed that the proposed method has the highest entropy content than its counterparts, thereby indicating the effectiveness of our algorithm in preserving the information content. In addition, the proposed method has the second best $Q_{AB/F}$ values for the various image frames.

In terms of the execution time, the proposed method takes around 38 seconds, whereas the conventional NSCT and the HOSVD approaches take around 224 and 542 seconds, respectively, to produce the fused outputs of dimension 512*512. This computational simplicity of the proposed algorithm is mainly due to the initial decomposition by DWT

Figure 4.3: Qualitative analysis: Source image [3]: A1 - Thermal, A2 - Visible (frame number: 25), Fused Results: [7] Laplacian (B1), [2] HOSVD fused (B2), [11] NSCT (C1) , Proposed method (C2)

.

Figure 4.4: Qualitative analysis: Source image [3]: A1 - Thermal, A2 - Visible (frame number: 50), Fused Results: [7] Laplacian (B1), [2] HOSVD fused (B2), [11] NSCT (C1) , Proposed method (C2)

.

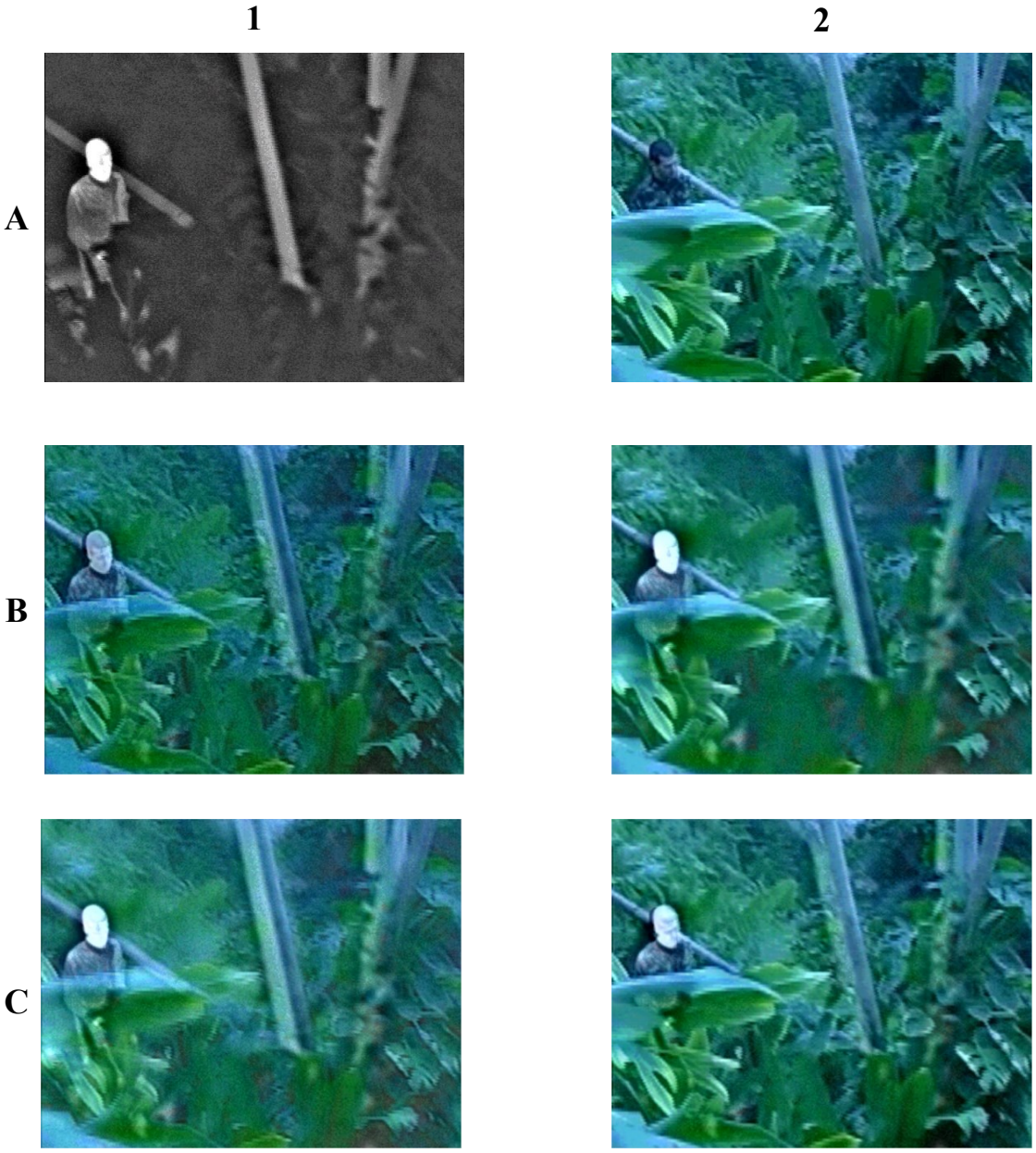

Figure 4.5: Qualitative analysis: Source image [3]: A1 - Thermal, A2 - Visible (frame number: 96), Fused Results: [7] Laplacian (B1), [2] HOSVD fused (B2), [11] NSCT (C1) , Proposed method (C2)

.

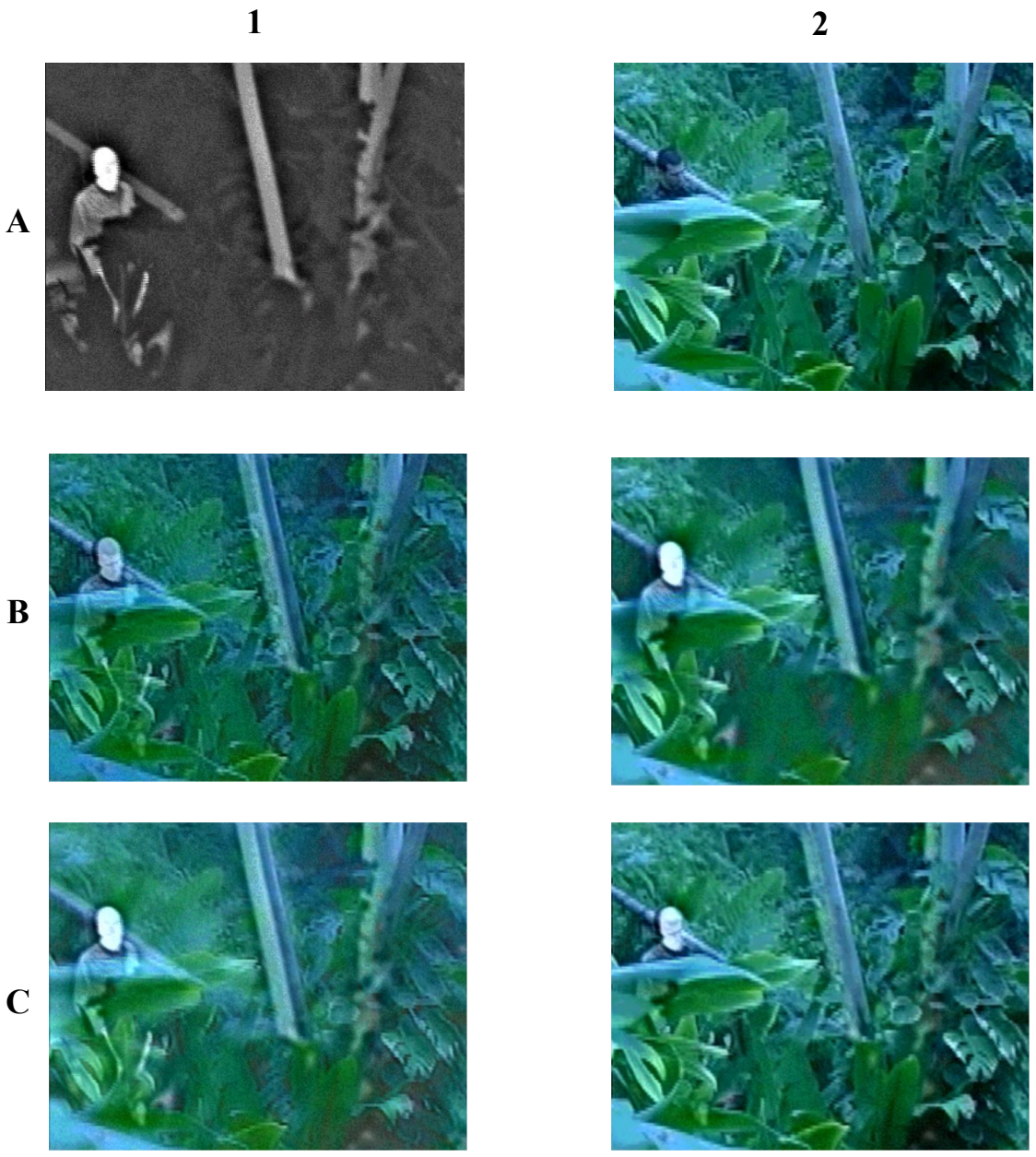

Figure 4.6: Qualitative analysis: Source image [3]: A1 - Thermal, A2 - Visible (frame number: 100), Fused Results: [7] Laplacian (B1), [2] HOSVD fused (B2), [11] NSCT (C1) , Proposed method (C2)
.

Figure 4.7: Fused Results using the target highlighting mode: frame number: 25 (A1), frame number: 50 (A2), frame number: 96 (B1), frame number: 100 (B2)
.

Table 4.1: Quantitative performance of the proposed video fusion algorithm

Frame Number of the Input videos	**$Q_{AB/F}$**				**Entropy**			
	[7]	**[2]**	**[11]**	**Proposed Method (Normal mode)**	**[7]**	**[2]**	**[11]**	**Proposed Method (Normal mode)**
25	0.4316	0.5257	0.3766	0.4624	6.7691	7.1236	0.3766	7.3158
50	0.4300	0.5247	0.3791	0.4609	6.7331	7.0700	7.0064	7.2600
96	0.4449	0.5229	0.4041	0.4336	6.8509	6.8509	6.8782	7.1844
100	0.4560	0.5192	0.4039	0.4419	6.7156	6.7156	6.8660	7.2114

Platform used: Matlab version R2012 a, **System specifications:** Intel Pentium P6200, 6GB DDR3 Memory.
Methods [7], [2] and [11] corresponds to Laplacian, HOSVD and conventional NSCT, respectively

and computationally efficient region-based fusion rules. However, the execution time of the proposed algorithm is slightly more than that of the Laplacian method, which is around 1.5 seconds. This is reasonable since our proposed method is based upon a combination of multi-scale transforms.

4.4 Summary

In this chapter, a novel video fusion algorithm capable of producing high-quality fused image frames even with a two-scale decomposition has been proposed. It has been shown that, in general, our approach is highly successful in exposing hidden camouflaged targets by preserving the features such as the thermal content and visible background from the thermal and visible videos. Furthermore, our algorithm is capable of producing high quality fused video in a very short span of time. The aforementioned merits qualify the proposed video fusion algorithm as an ideal candidate for real-time detection of camouflaged targets.

Chapter 5

Conclusion

In this thesis, we have considered the fusion of digital images and videos using a multi-scale transform based fusion framework. First, we have proposed a novel low-complexity framework for the fusion of multi-modal images based on an improved multi-scale decomposition framework and pixel-based fusion rules.

The main features of the proposed image fusion algorithm are summarised below.

1. The proposed framework uses a combination of non-subsampled contourlet and wavelet transforms for the initial multi-scale decompositions.

2. The decomposed multi-scale coefficients are then fused twice using various computationally inexpensive local activity measures.

3. Experimental results have shown that the proposed approach performs better or on par with the existing state-of-the art image fusion algorithms in terms of the quantitative and qualitative results, thereby making it an ideal candidate for a wide variety of applications such as remote sensing, medical imaging and computer vision.

4. In addition, the proposed image fusion algorithm can produce high quality fused images even with a computationally inexpensive two-scale decomposition.

The primary characteristics of the proposed video fusion algorithm are summarized below.

1. The proposed video fusion framework is highly successful in retaining the thermal content as well as the visible-background content, which is evident from the quantitative and qualitative results.

2. The proposed video fusion algorithm when used in the target highlighting mode further enhances the hidden target, thereby making the camouflaged targets highly visible and easier to track.

3. Experimental results have shown that the proposed video fusion approach outperforms its existing counterparts by a very large margin in terms of the execution time, thereby making it an ideal candidate for real-time multi-modal video surveillance applications.

4. In addition, the proposed video fusion algorithm can produce high quality fused video even with a relatively low two-scale decomposition.

5.1 Future Work

The complexity of the proposed algorithms is relatively low, when compared to its counterparts such as conventional NSCT [11] and HOSVD [2]. But, the execution time of the proposed algorithms is still slightly higher than that of the Laplacian method [7]. This is mainly due to the multi-scale decomposition of the NSCT involved in our proposed algorithms. In MATLAB, for example, the proposed video fusion algorithm takes around 32 seconds to produce the fused output for a 512*512 image frame . This can be substantially reduced by implementing the proposed algorithms using a programming language, such as C/C++. This would drastically reduce the multi-scale decomposition time of the NSCT involved in our algorithms, thereby further increasing its claim for real-time image and video fusion applications. This claim, of course, needs to be confirmed only after implementing the proposed algorithms in a high-level language. Even though the processing time of the proposed algorithms in MATLAB is almost close to real-time, a hardware implementation of the proposed algorithms could be another area that should be taken

up in the future after the initial implementation using C/C++. This hardware implementation can be further used as an integrated component for real-time video surveillance in paramilitary applications, which is based on the fusion of inputs from various multi-modal sensors such as thermal and visible cameras.

References

[1] G. Bhatnagar, Q. M. J. Wu, and L. Zheng, "Directive contrast based multimodal medical image fusion in nsct domain," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1014–1024, Aug. 2013.

[2] J. Liang, Y. He, D. Liu, and X. Zeng, "Image fusion using higher order singular value decomposition," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2898–2909, May. 2012.

[3] [Online]. Available http://www.imagefusion.org.

[4] A. A. Goshtasby and S. Nikolov, "Image fusion: Advances in the state of the art," *Information Fusion*, vol. 8, no. 2, pp. 114 – 118, Apr. 2007.

[5] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Information Fusion*, vol. 12, no. 2, pp. 74 – 84, Apr. 2011.

[6] L. Jacques, L. Duval, C. Chaux, and G. Peyré, "A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity," *Signal Processing*, vol. 91, no. 12, pp. 2699 – 2730, Dec. 2011.

[7] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. Qgden, "Pyramid methods in signal processing," *RCA Eng*, vol. 29, no. 6, pp. 33–41, Nov./Dec. 1984.

[8] K. Amolins, Y. Zhang, and P. Dare, "Wavelet based image fusion techniques an introduction, review and comparison," *Journal of Photogrammetry and Remote Sensing*, vol. 62, no. 4, pp. 249 – 263, Jul. 2007.

[9] A. L. da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.

[10] D. W. Townsend and S. R. Cherry, "Combining anatomy and function: the path to true image fusion," *European radiology*, vol. 11, no. 10, pp. 1968–1974, Jul. 2001.

[11] X. B. Qu, W. J. Yan, H. Z. Xiao, and Z. Q. Zhu, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Automatica Sinica*, vol. 34, no. 12, pp. 1508 – 1514, Dec. 2008.

[12] M. Yazdi and E. K. Ghasrodashti, "Image fusion based on non-subsampled contourlet transform and phase congruency," in *Proceedings of the 19th International Conference on Systems, Signals and Image Processing*, pp. 616–620, Apr. 2012.

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[14] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing*, vol. 29, no. 3, pp. 273 – 285, Mar. 1985.

[15] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

[16] G. Piella, "A region-based multiresolution image fusion algorithm," in *Proceedings of the 5th International Conference on Information Fusion*, pp. 1557–1564, Jul. 2002.

[17] C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, Feb. 2000.

[18] S. S. Pillai and M. N. S. Swamy, "A fast, low complexity image fusion algorithm based on multiscale transforms," in *Proceedings of the 56th International Midwest Symposium on Circuits and Systems*, pp. 1286–1289, Aug. 2013.

[19] W. Burger and M. J. Burge, *Principles of Digital image processing: Fundamental Techniques*. Springer, 2009.

[20] P. J. Burt and R. J. Kolczynski, "Enhanced image capture through fusion," in *Proceedings of the 4th International Conference on Computer Vision*, pp. 173–182, May 1993.

[21] T. Stathaki, *Image Fusion: Algorithms and Applications*. Academic press, 2008.

[22] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 493 – 500, Mar. 2007.

[23] [Online]. Available http://www.mathworks.com/matlabcentral/fileexchange/10049.

[24] [Online]. Available http://www.metapix.de/fusetool.zip.

[25] [Online]. Available http://dspace.xmu.edu.cn:8080/dspace/handle/2288/8332.

[26] [Online]. Available http://xudongkang.weebly.com.

[27] [Online]. Available http://en.wikipedia.org/wiki/YUV.