CRANFIELD UNIVERSITY

WASNAA WITWIT

WAVELET-BASED IMAGE AND VIDEO SUPER-RESOLUTION RECONSTRUCTION

SCHOOL OF AROSPACE TRANPORT AND MANUFACTURING
PhD in Manufacturing

PhD
Academic Year: 2014 - 2017

Supervisors: Dr. Yifan Zhao and Dr. Karl Jenkins
December 2017

CRANFIELD UNIVERSITY

SCHOOL OF AROSPACE TRANSPORT AND MANUFACTURING
PhD in Manufacturing

PhD

Academic Year 2014 - 2017

WASNAA WITWIT

WAVELET-BASED IMAGE AND VIDEO SUPER-RESOLUTION RECONSTRUCTION

Supervisors: Dr. Yifan Zhao and Dr. Karl Jenkins
December 2017

This thesis is submitted in partial fulfilment of the requirements for the degree of PhD
(NB. This section can be removed if the award of the degree is based solely on examination of the thesis)

ABSTRACT

Super-resolution reconstruction process offers the solution to overcome the high-cost and inherent resolution limitations of current imaging systems. The wavelet transform is a powerful tool for super-resolution reconstruction. This research provides a detailed study of the wavelet-based super-resolution reconstruction process, and wavelet-based resolution enhancement process (with which it is closely associated). It was addressed to handle an explicit need for a robust wavelet-based method that guarantees efficient utilisation of the SR reconstruction problem in the wavelet-domain, which will lead to a consistent solution of this problem and improved performance.

This research proposes a novel performance assessment approach to improve the performance of the existing wavelet-based image resolution enhancement techniques. The novel approach is based on identifying the factors that effectively influence on the performance of these techniques, and designing a novel optimal factor analysis (OFA) algorithm. A new wavelet-based image resolution enhancement method, based on discrete wavelet transform and new-edge directed interpolation (DWT-NEDI), and an adaptive thresholding process, has been developed. The DWT-NEDI algorithm aims to correct the geometric errors and remove the noise for degraded satellite images. A robust wavelet-based video super-resolution technique, based on global motion is developed by combining the DWT-NEDI method, with super-resolution reconstruction methods, in order to increase the spatial-resolution and remove the noise and aliasing artefacts. A new video super-resolution framework is designed using an adaptive local motion decomposition and wavelet transform reconstruction (ALMD-WTR). This is to address the challenge of the super-resolution problem for the real-world video sequences containing complex local motions.

The results show that OFA approach improves the performance of the selected wavelet-based methods. The DWT-NEDI algorithm outperforms the state-of-the art wavelet-based algorithms. The global motion-based algorithm has the best performance over the super-resolution techniques, namely Keren and structure-adaptive normalised convolution methods. ALMD-WTR framework surpass the state-of-the-art wavelet-based algorithm, namely local motion-based video super-resolution.

Keywords:

Resolution enhancement; discrete wavelet transform; interpolation; registration; global motion; local motion; satellite images

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor Dr Yifan Zhao for his professional guidance and valuable suggestions during the planning and development of this research. His willingness to give his time so generously has been very much appreciated. I would like to express my great appreciation to my supervisor Dr Karl Jenkins for his valuable support and constructive suggestions on this project. I would like to give my special thanks for the Ministry of Higher Education and Scientific Research of Iraq for offering me the funding to running this research.

My special thanks are extended to my dad Jaafar Witwit for his patient guidance and encouragement to do this research. I am particularly grateful for my mum, my sisters Mayassa, Refref, Aumnia and my brother Ali who support and encourage me throughout my study.

TABLE OF CONTENTS

LIST OF PUBLICATION

1. Wasnaa Witwit, Yitian Zhao, Karl Jenkins, and Yifan Zhao, "An Optimal Factor Analysis Approach to Improve the Wavelet-based Image Resolution Enhancement Techniques," Global Journal of Computer Science and Technology: F Graphic and Vision, vol. 16, no. 3, pp. 11-20, 2016, IF:1.400.
2. Wasnaa Witwit, Yifan Zhao, Karl Jenkins, and Yitian Zhao, "Satellite Image Resolution Enhancement using Discrete Wavelet Transform and New Edge-Directed Interpolation," Journal of Electronic Imaging, vol. 26, no. 2, pp. 1-9, 2017, IF:0.616.
3. Wasnaa Witwit, Yifan Zhao, Karl Jenkins, and Sri Addepalli, "Global Motion based Video Super-Resolution Reconstruction using Discrete Wavelet Transform," Journal of Multimedia Tools and Applications, IF:1.530, Accepted on 26/03/2018.
4. Wasnaa Witwit and Yifan Zhao, "Video Super-Resolution based on Adaptive Local Motion Decomposition and Wavelet Transform Reconstruction," Journal of Neurocomputing, IF:3.137, to be submitted.

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

HR	High-Resolution
SR	Super-Resolution
LR	Low-Resolution
CCD	Charge-Coupled Device
CMOS	Complementary Metal-Oxide-Semiconductor
WT	Wavelet Transform
DWT	Discrete Wavelet Transform
OFA	Optimal Factor Analysis
NEDI	New Edge-Directed Interpolation
DWT-NEDI	Discrete Wavelet Transform-New Edge Directed Interpolation
ALMD-WTR	Adaptive Local Motion Decomposition-Wavelet Transform Reconstruction
IBP	Iterative Back-Projection
POCS	Projection Onto Convex Sets
ML	Maximum Likelihood
MAP	Maximum A Posteriori
CWT	Complex Wavelet Transform
SWT	Stationary Wavelet Transform
PSF	Point-Spread Function
SANC	Structure-Adaptive Normalized Convolution
NC	Normalized Convolution
CAT	Computer Aided Tomography
GUI	Graphical User Interface
EM	Expectation Maximization
PDF	Probability Density Function
CG	Conjugate Gradient
GCV	Generalized Cross-Validation
WZP	Wavelet Zero Padding
IWT	Inverse Wavelet Transform
HMT	Hidden Markov Tree
HMM	Hidden Markov Models
CS	Cycle-Spinning

DT-CWT	Dual-Tree Complex Wavelet Transform
LL	Low-Low
LH	Low-High
HL	High-Low
HH	High-High
DASR	Demirrel-Anbarjafari Super-Resolution
EDI	Edge-directed Interpolation
FT	Fourier Transform
SBS	Simple Background Subtraction
RA	Running Average
SDE	Sigma-Delta Estimation
NLM	Non-Local Means
SKR	Steering Kernel Regression
ANR	Anchored Neighbour Regression
SRHDT	Super-Resolution Hieracrchical Decision Trees
SRCNN	Super-Resolution Convolution Neural Network
MAR	Multi-Resolution Analysis
DB	Daubechies
SYM	Symlets
COIF	Coiflets
BIOR	Biorthogonal
WZP-CS	Non-Directional Cycle Spinning
WZP-DCS	Directional Cycle Spinning
RMSE	Root-Mean-Square Error
PSNR	Peak-Signal-to-Noise Ratio
MSE	Mean-Square-Error
MI	Multi-Input
MO	Multi-Output
FOM	Figure Of Merit
SSIM	Structure Similarity
E	Entropy
PM	Pixel-based Motion Mask
BM	Block-based Motion Mask

1 Introduction

1.1 Research Background

High-resolution (HR) images and videos are highly desirable, and in strong demand for most electronic imaging applications: they are used to both providing better content visualisation, and for extracting additional detailed information. However, it is difficult to obtain the requisite HR images/videos in many practical applications. This is due to the high cost and inherent physical constraints of the precision optics and imaging sensors involved, or the limitations in bandwidth of data communications. Therefore, *super-resolution (SR)* reconstruction has emerged as an alternative low-cost solution, and is attracting increasing attention. *Super-resolution* reconstruction aims to reconstruct a single HR image, or a set of HR images (video), from a sequence of different low-resolution (LR) images captured from the same scene in order to overcome the limitations and/or possibly ill-posed conditions of the imaging system [1], [2]. SR reconstruction has been a very active area of research over the last two decades for many applications in various fields, such as satellite imaging [3], astronomical observation [4], medical diagnostics [5], forensic imaging [6], video surveillance systems [7], and other applications [8].

A HR image means that the number of pixels per unit area is high, and thus the HR image contains more detail; this can be crucial in particular applications [1]. Image resolution is classified into five different types: pixel resolution, spatial resolution, spectral resolution, temporal resolution, and radiometric resolution. *Spatial resolution* is the primary focus of this research. *Spatial resolution* refers to the number of pixels per unit area contained in an image, and it is measured in pixels per unit area. The spatial resolution of a digital image is initially determined by the imaging sensors, or the imaging systems. A modern image sensor is typically either a charge-coupled device (CCD), or a complementary metal-oxide-semiconductor (CMOS) active-pixel sensor. These sensors consist of a two-dimensional array of photo-detector elements or pixels. The higher the number of detector elements per unit area, the higher the spatial resolution possible of the image acquisition device. An imaging system with an insufficient number of detectors will produce a LR image with an aliasing effect; when a scene is captured with a LR device, it is sampled at a low sampling rate, causing the aliasing effect. To increase the spatial resolution of an imaging system, there are two possible hardware-based solutions [9], [10], [11].

A straightforward solution to increase spatial resolution is to reduce the pixel size with enhanced sensor manufacturing techniques. However, as the pixel size decreases, the amount of light incident on that pixel also decreases; this causes shot noise, severely degrading the quality of the image. Furthermore, the hardware cost of the sensor will also increase with the increase in pixel density. Reducing the pixel size without causing the shot noise is subject to physical constraints; the optimal pixel size is estimated at about $40um^2$ for a CMOS sensor. The current imaging sensor technology has almost reached this limitation [1], [11], [12].

Another solution to enhance spatial resolution is to increase the chip size, which results in an increase in capacitance. This means more pixels will be involved in an imaging system. However, this solution is not considered to be effective; a large capacitance slows down the charge transfer rate, leading to a longer period of time to acquire an image. Moreover, the hardware cost of an image sensor will also increase with the increase in chip size. Therefore, it is essential to find an effective image processing technique to increase the spatial resolution level for a low-cost, without replacing the existing imaging system. To address this challenge, the concept of super-resolution has been developed [1], [11], [12].

The fundamental notion behind SR reconstruction for increasing spatial resolution relies on the availability of a sequence of LR images acquired from the same scene. That is, the LR images represent different views of the same scene with sub-pixel shifts, and thus each LR image contains additional information introduced by sub-pixel shifts between the LR images. If the LR images are shifted by integer pixel shifts, each LR image contains the same information, and therefore there is no extra information that can be combined to generate a HR image. However, if the LR images are shifted by sub-pixel shifts and down-sampled (aliased), each LR image contains additional information that can be exploited to reconstruct a HR image.

To obtain different viewpoints (multi-view analysis) of the same scene, relative scene motions (sub-pixel shifts) must exist between the LR images from multiple scenes or video sequences. Multiple scenes can be acquired by using a single camera with successive captures, or by employing multiple cameras positioned in different locations. These scene motions can occur due to controlled motions in imaging systems (e.g. images captured from orbiting satellites); or due to uncontrolled motions between the imaging system and scene (e.g. movement of local objects or vibration of imaging systems) [11], [12]. If these sub-pixel shifts are known or can be estimated with sub-pixel accuracy, SR reconstruction is possible by combining these LR images, as illustrated in Figure 1-1.

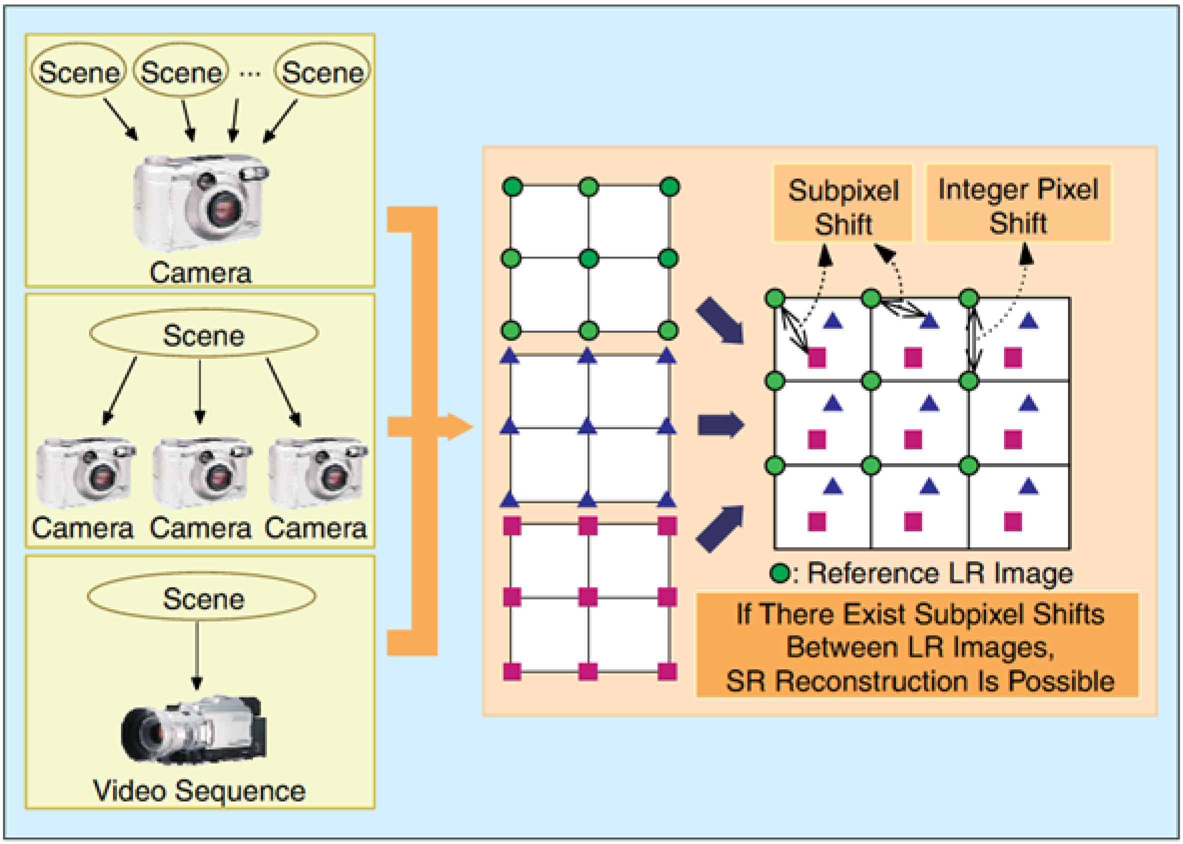


Figure 1-1 The basic premise for super-resolution [1]

SR image/video reconstruction is a severely ill-posed inverse problem because of an inadequate number of LR images, ill-conditioned motion estimation/registration, and unknown degradation operators. There are two major challenges in SR reconstruction: image registration and image reconstruction. *Image registration* is the process of aligning two or more images (the reference and observed images) captured from the same scene at different times, from different viewpoints, and/or by different sensors [13]. Registration/motion estimation aims at estimating the motion between the reference LR image and each of the neighbouring LR images at sub-pixel accuracy. It plays a critical role in the success of SR reconstruction, and it is widely understood as an ill-posed estimation problem. The problem becomes more difficult in building a practical SR system; the observed LR images contain large amounts of aliasing artefacts. The performance of the basic registration algorithms degrades rapidly when the resolution of the observed images goes down, leading to more registration errors [11], [14].

Image reconstruction aims to combine a sequence of different registered images into a single HR image, and eliminate any blur and noise that occurred during image acquisition [15]. Most SR reconstruction approaches can be classified into either frequency-domain-based or spatial-domain-based approaches. Between these two domains, the wavelet-domain-based SR approach has more recently emerged as an effective tool for many image and signal processing applications. The wavelet transform-based approach is able to provide both frequency information and temporal information in the transformation process. The attractive properties of the wavelet transform (WT), such as locality, multi-resolution, and compression enable it to analyse real-world signals [16]. The wavelet transform is a tool that decomposes an image into low and high-frequency sub-bands, and then examines each sub-band with a resolution matched to its scale [17]. The advantage of wavelet decomposition is that image characteristics at different scales can be isolated and analysed: global characteristics can be examined at coarse scales, while local characteristics can be analysed at fine scales [18]. The essential idea behind wavelet-based SR reconstruction approaches is the estimation of unknown wavelet coefficients in sub-bands containing high-pass frequency spatial information, in order to reconstruct a HR image from the observed LR images. Discrete wavelet transform (DWT) is one of the recent wavelet transforms: it being used as a powerful tool for resolution enhancement, and in many image and video processing applications [19]. The DWT provides a sufficient quantity of information for local analysis and synthesis of an image [12].

A closely related approach to SR reconstruction is image interpolation, which can also be used to increase the number of pixels in an image. *Image interpolation* aims to estimate values at unknown locations using known pixel constraints [19]. Interpolation is extensively used in several image processing applications, including region-of-interest image magnification, image rotation, and sub-pixel image registration. However, as there is no extra information introduced, the quality of the interpolated image is inherently limited due to the use of a single, aliased LR image. Another limiting factor is the number of pixel constraints present within the data [11], [12]; the interpolated image does not reconstruct the missing high-frequency information, and does not remove the aliasing effect [14]. Consequently, multi-frame SR reconstruction approaches have been proposed to exploit the additional information provided by a sequence of temporarily correlated LR frames with sub-pixel accuracy in order to improve the resolution. Another related approach to SR reconstruction, that cannot be used to change the size of the image, is image restoration. Image restoration endeavours to recover an original HR image from a degraded image by applying a priori knowledge of the degradation operators and inverse process.

1.2 Research Motivation

SR reconstruction technology offers a means to overcome the high-cost and the inherent physical resolution limitations of hardware-based approaches, for many practical applications. For instance, satellite imaging is one such an application as updating the cameras in a satellite is hugely expensive. Satellites, e.g. Landsat and WorldView-2, can acquire multi-temporal or multi-view images of the same area and thus provide an opportunity for SR. A significant example of

SR is shown in [3] for remote sensing images provided by multi-angle WorldView-2, which incorporate five angular images. The main challenge for remotely sensing image SR is to tackle the scene changes caused by temporal differences. In astronomical observation, the physical resolution limitations of astronomical devices also provide the possibility for SR technologies to play a key role. Therefore, SR techniques have been developed to help astronomers with the exploration the outer space by improving the spatial resolution of small objects. A specific example of SR is given in [4] for multiple star images. In medical diagnosis, resolution limitations of medical imaging systems usually degrade the quality of images in the diagnosis. Thus, SR techniques have been used to increase the spatial resolution of medical images while preserving the true isotropic 3-D imaging. In surveillance, digital video recorder devices play a significant role in applications, such as traffic surveillance and security monitoring. However, it is difficult to supply large-scale HR devices. This is because outdoor video devices are vulnerable to weather conditions. Therefore, the need for video SR technologies is necessary [8]. These applications make SR techniques a particularly requisite and relevant research topic in our daily life.

Most SR reconstruction approaches employ either the spatial-domain, or frequency-domain, to reconstruct a HR image from given LR images. The wavelet-domain-based SR reconstruction approach is able to employ both the spatial and frequency-domains, and integrate properties of both to reconstruct a HR image from observed LR images. The wavelet-domain-based SR has not yet been fully exploited in SR reconstruction. This work is a detailed study of the wavelet-based SR problem, and wavelet-based resolution enhancement problem (with which it is closely associated). This research identifies previous methods, their associated limitations, and the major differences between them. This work then develops new algorithms that solve these problems more efficiently. This work also focused on the study of SR reconstruction in the spatial-domain and frequency-domain. This research clarifies the merits of wavelet-domain in terms of the conceptual properties, associated strategies and achieved challenges. There is an explicit need for a robustness wavelet-based method that guarantees efficient utilisation of the SR reconstruction problem in the wavelet-domain, and considers different scenarios of this problem, which will lead to a consistent solution of this problem and improved performance. As a result, there has an explicit lack of a generally accepted and validated definition of this problem which has lead, in turn, to inconsistent solutions and limited performance.

This research will

1. Identify the SR reconstruction problem in different domains (spatial, frequency, and wavelet).
2. Address the wavelet-based SR reconstruction problem and wavelet-based resolution enhancement problem.
3. Develop a robust algorithm for wavelet-based SR reconstruction demonstrating an improved performance.

1.3 Research Questions

The main research question of this study is to investigate how wavelet-based approach can be used to improve the SR performance and go beyond the resolution limit of the LR observations, rather than improving the expensive hardware devices. In order to form the framework for the required work, the following research questions were formulated:

1. What is the current performance bottleneck of the existing wavelet-based image resolution enhancement methods?
2. Why these methods use inconsistent factors to assess the performance and what is the impacts of using different factors on each method?
3. How to address resolution enhancement problem in the wavelet-domain and increase the spatial resolution for a single observed noisy LR image and a particularly more challenging satellite images?
4. How to compensate for the errors and remove potential noise caused by the limitations of imaging systems?
5. How to solve SR reconstruction problem in the wavelet-domain and increase the spatial resolution for the observed noisy video sequences captured from a moving camera which feature global motion?

6. What is the current performance limitations of the existing wavelet-based SR methods working under global motion and how to investigate the effectiveness of the parameters that influence on the SR performance?
7. How to solve SR reconstruction problem in the wavelet-domain and increase the spatial resolution for real-world video sequences which feature complex local motions caused by multiple moving objects?
8. What is the current performance drawbacks of the existing wavelet-based SR methods working under complex local motions and how to overcome local motion inaccuracies?

1.4 Research Aim and Objectives

The attractive primary properties of the wavelet transform, such as locality and multi-resolution make it effective for analysing real-world signals at more than one resolution. The wavelet-based SR reconstruction is a new research area, which has recently received increasing attention in the digital image processing. The primary aim of this project was to design and develop a set of wavelet-based reconstruction approaches in order to increase spatial-resolution and overcome the inherent physical resolution limitations for different imaging systems.

In order to accomplish the aim of this study, the following four objectives were defined:

1. To design a novel performance assessment approach for improving the existing wavelet-based image resolution enhancement techniques.
2. To develop a new wavelet-based image resolution enhancement method for increasing the spatial resolution of a single LR image.
3. To develop a robust wavelet-based multi-frame (video) super-resolution method for increasing the spatial resolution when the camera is moving and the observed object is static.
4. To design a new wavelet-based video super-resolution framework for increasing the spatial resolution when the camera is stationary and the observed object is moving.

The reason for designing a novel performance assessment approach to improve the performance of the existing wavelet-based methods is that these methods make a number of assumptions with regard to the factors that affect the performance and they use inconsistent factors to assess the performance, which limit their application in various fields. Therefore, the need for a more advanced approach to find the best factors so as to increase the applicability and improve the overall performance of the existing wavelet-based methods. Although many wavelet-based image resolution enhancement methods have been developed progressively in various fields for different practical applications, resolution enhancement of satellite images still remains a challenge. This is because satellite imaging is one such an application as improving the sensors on satellites is the most cost means. In addition, the applicability of these methods to satellite images is still limited by the relatively poor performance and time consuming. Thus, the demand for developing a more efficient wavelet-based image resolution method to improve the performance is highly desirable.

The justification for proposing a new SR reconstruction technique in the wavelet-domain to increase the spatial resolution for video sequences with global motion is that the application of the proposed technique is especially useful when the camera is moving and the observed object is stationary, and the existing wavelet-based SR techniques have limited performance capabilities for the parameters that influence on the SR performance. Therefore, the need for proposing a robust SR algorithm to provide high performance and improved flexibility for these various parameters. Although, the application of the proposed algorithm is very useful for video sequences with global motion, dealing with real-world video sequences containing complex local motions is still a challenge. This is because some wavelet-based SR algorithms can deal with local motions, but the local motion inaccuracies caused by multiple moving objects, and computational cost limit their performance in different application fields. Thus, the demand for designing a new SR reconstruction technique in the wavelet-domain to reduce the local motion errors and increase the spatial resolution is still highly requested in reduced computation intensive.

1.5 Research Methodology

The research methodology of this Thesis is illustrated by Figure 1-2. This PhD research starts from reviewing the classic and state-of-the-art SR reconstruction approaches in both spatial and frequency-domains, with a focus on wavelet-domain-based SR approaches through the literature review, and then identifying the research gap. A sensitivity analysis of various factors of the existing wavelet-based methods is conducted to determine if there is a prospect to improve the performance of these methods in a relatively straightforward way, which is the first major objective of this research. In this objective, the performance evaluation measures are reviewed and improved, which may help select the optimal factors. Such a study is important and essential because it contributes to the development of new wavelet-based approaches through guiding the selection of optimal factors.

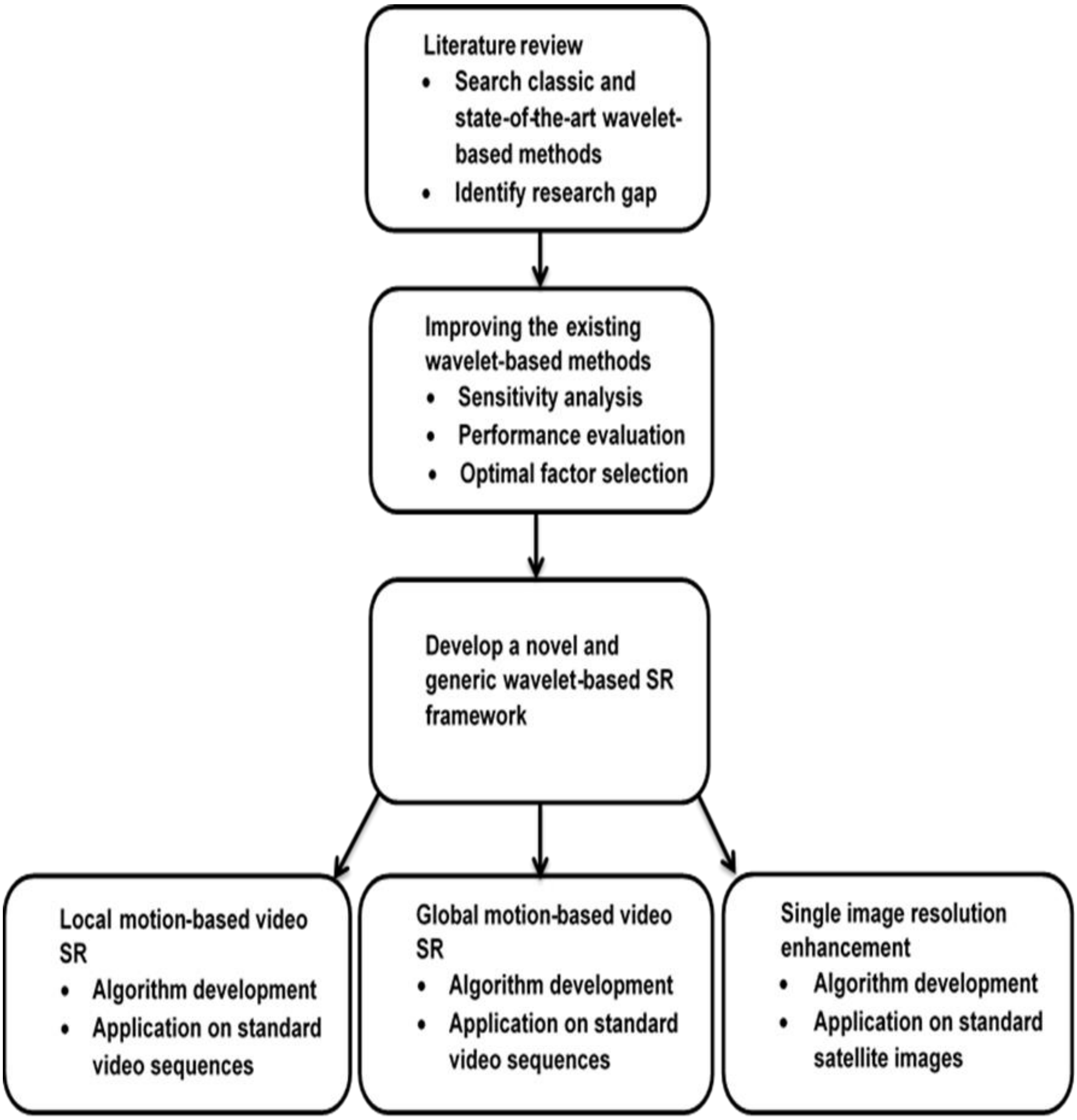


Figure 1-2 Research Methodology

The research is then to develop a novel and generic wavelet-based SR reconstruction framework, which is the main contribution of knowledge of this Thesis. Following the "simple to complex" strategy, the new framework is then tailored on various applications, starting from single image resolution enhancement, global motion-based video SR to local motion-based video SR, which are the next three major objectives respectively. The proposed new framework aims to have wide applications on various imaging scenarios. It should be noted that the algorithm for each application is tailored considerably due to natural difference of imaging systems, although the way to employ the wavelet transform is similar. The research methodology of each objective is summarised below:

1.5.1 Designing Performance Assessment Approach

The first objective is devoted to a study of the existing wavelet-based image resolution enhancement techniques with regard to the way to assess their performance. This objective focuses on identifying the assumptions regarding the factors that influence on the performance of these techniques, and investigating the inconsistent factors used for each technique, which limit their performance in practice. The first step contributes a sensitivity analysis of the important factors and evaluating the impacts of the prevailing assumptions, see Chapter 4 for more details.

The output of this step helps the development of a new wavelet-based approach to guide the selection of optimal factors in a more comprehensive approach, in order to increase the applicability and better improve the performance of the current wavelet-based methods. The proposed algorithm is based on an exhausted search in finding the best parameters, a detailed mathematical model formulation and solution is discussed in Chapter 4. A new metric is also proposed to better measure the overall performance. This objective provides a better understanding of the resolution enhancement problem in the wavelet-domain, plays a significant role to connect between the other objectives, and reveals what is the current performance bottleneck of the existing wavelet-based image resolution enhancement methods.

In terms of the used testing samples, this objective uses three well known standard images (Lena, Baboon, and Elaine) with different features as test images to assess the performance of the considered wavelet-based resolution enhancement methods.

1.5.2 Developing Single Image Resolution Enhancement Method

This objective is to develop a new wavelet-based image resolution enhancement approach for addressing the resolution enhancement problem of the more challenging degraded satellite images. This approach focuses on satellite images acquired by cameras on satellites, which can have errors and are most corrupted by noise. The developed method is based on integrating properties from both the frequency-domain and spatial-domain by connection between concepts of wavelet-based interpolation and wavelet-based de-noising in order to increase the spatial resolution. This objective aims to provide a more efficient algorithm, which has low computational expensive and time consuming.

The proposed method is based on using the interpolation of separated high-frequency sub-bands generated by wavelet transform in order to preserve more edges and correct the errors. The developed algorithm is also based on employing the wavelet thresholding of estimated high-frequency sub-bands in order to boost the edges and eliminate the potential nose. The concept of wavelet thresholding process is that the signal energy is often preserved on a few coefficients whereas the noise energy is spread in all coefficients in the wavelet-domain. The proposed approach aims to keep few large coefficients representing the signal whilst suppressing noise coefficients to zero. The details about the chosen methods and the hypothesis of application these methods are discussed in Chapter 5.

The developed method is tested firstly on the same standard images used in Chapter 4, and then it is tested on the real satellite images in order to validate and compare the variation in performance in terms of two types of different images, see Section 5.4 Results and Discussion for more details. The performance of the proposed method is also compared with the interpolation methods and state-of-the-art wavelet-based resolution enhancement methods.

1.5.3 Developing Global motion Video Super-Resolution Method

The third objective is dedicated to proposing a robust wavelet-based SR technique for solving the SR reconstruction problem of the observed noisy video sequences containing global motion. The technique proposed is based on combining the developed single wavelet-based resolution enhancement approach, with SR reconstruction methods. The purpose of the technique developed is to integrate merits from these methods in both the spatial and frequency-domains for increasing the spatial resolution, while supressing the noise and aliasing artefacts.

Another purpose of the algorithm developed is to address the limited performance of most the existing wavelet-based SR methods for the parameters that influence on the SR performance, and investigate the sensitivity analysis of the proposed algorithm for these parameters. This task contributes determining the parameters that affect the SR performance, analysing how these parameters can affect the performance of the implemented technique, and discussing the effectiveness of each parameter. This objective provides improved performance and flexibility of the proposed technique over the existing wavelet-based SR techniques working under global motion, see Chapter 6 for more details.

To validate and evaluate the performance of the proposed SR technique, three well-known video sequences, namely, "Akiyo", "Mother & daughter", and "Forman" are tested. The performance of the proposed technique is also evaluated in comparison with interpolation techniques, and state-of-the-art wavelet-based resolution enhancement and global motion-based video SR techniques, see Section 6.4 Results and Discussion for more details.

1.5.4 Designing Local motion Video Super-Resolution Framework

This objective is to propose a new SR framework for solving the SR problem of the challenge practical video sequences with complex local motion between video frames. The proposed method is based on dividing the observed LR frame into multiple types of blocks and regions based on motion characteristics, and then proposing wavelet-based SR approach or wavelet-based resolution enhancement approach for each type of blocks/regions. This method aims to increase the spatial resolution through reducing the local motion errors by adapting motion decomposition, and reducing the boundary artefacts. The details about the chosen methods are discussed in Chapter 7.

The proposed SR method is tested on four benchmark video sequences, namely, "Akiyo", "Mother & daughter", "Foreman", and "Ice" in order to evaluate the performance. The performance of the proposed method is also evaluated and validated in comparison with the interpolation methods, and state-of-the-art resolution enhancement and local motion-based video SR methods, see section 7.4 Results and Discussion for more details.

1.6 Thesis Structure

This thesis is divided into eight chapters, along with the references. The outline of structure is shown in Figure 1-2, and is summarised below:

Chapter 1 provides an introduction and a general overview of SR reconstruction as the solution to overcome the high-cost and inherent resolution constraints of current imaging systems: solving this for many practical applications. However, it is an ill-posed problem because of the major challenges in the two processes known as image registration and image reconstruction. This chapter also discusses the research motivation, the research questions, the research aim and objectives, and the research methodology of each objective.

Chapter 2 provides a review of the literature. The literature was examined according to a variety factors in order to classify SR reconstruction approaches, and the review covers all aspects related to wavelet-based SR reconstruction approaches. It starts with a taxonomy of the surveyed SR reconstruction methods in terms of the domain deployed; this survey consists of three categories: spatial-domain, frequency-domain, and wavelet-domain. A range of previous spatial-domain and frequency-domain SR algorithms are reviewed in this chapter in addition to their advantages and disadvantages. Finally, the SR taxonomy ends with a comprehensive literature review of predefined methods in the wavelet-based SR reconstruction. The literature is divided into two classes: the first includes the existing wavelet-based image resolution enhancement methods of single image; the second includes the existing wavelet-based SR methods of multiple images (video).

Chapter 3 presents the basic concepts of wavelet-based image resolution enhancement and SR reconstruction methods. This chapter contains the key elements for analysing the resolution enhancement and SR reconstruction problems in the wavelet-domain, which include the image observation model, the wavelet transform, wavelet bases, and types of wavelet transforms.

Chapter 4 presents a study of the predefined wavelet-based image resolution enhancement methods and a comparison of these methods based on the method of assessing their performance. This study was conducted in this research because the current methods to assess performance have many limitations with regard to the factors that influence on the performance which results in an inaccurate conclusion. The proposed solution is based on determining the factors that significantly affect the performance of wavelet-based methods, and designing a novel optimal factor analysis (OFA) algorithm for improving the performance of these methods. In this chapter, a new figure of merit measure is proposed to assist the selection of factors. The content of Chapter 4 has been published (see Publication 1 in the Publication List).

Chapter 5 discusses the development of a new wavelet-based image resolution enhancement algorithm, based on discrete wavelet transform (DWT) and new edge-directed interpolation (NEDI), to correct the geometric errors and improve the spatial resolution for low-quality satellite images degraded by geometric distortion. An adaptive thresholding process is proposed to preserve the true edges whilst removing the noise. The DWT-NEDI algorithm is based on the observations made in Chapter 4. The performance of the DWT-NEDI method is validated with regard to different classes of satellite images, and in terms of a comparison with the state-of-the-art wavelet-based resolution enhancement methods. In this chapter, four new image quality measures are proposed to better compare the overall performance for multiple images. The content of Chapter 5 has been published (see Publication 2 in the Publication List).

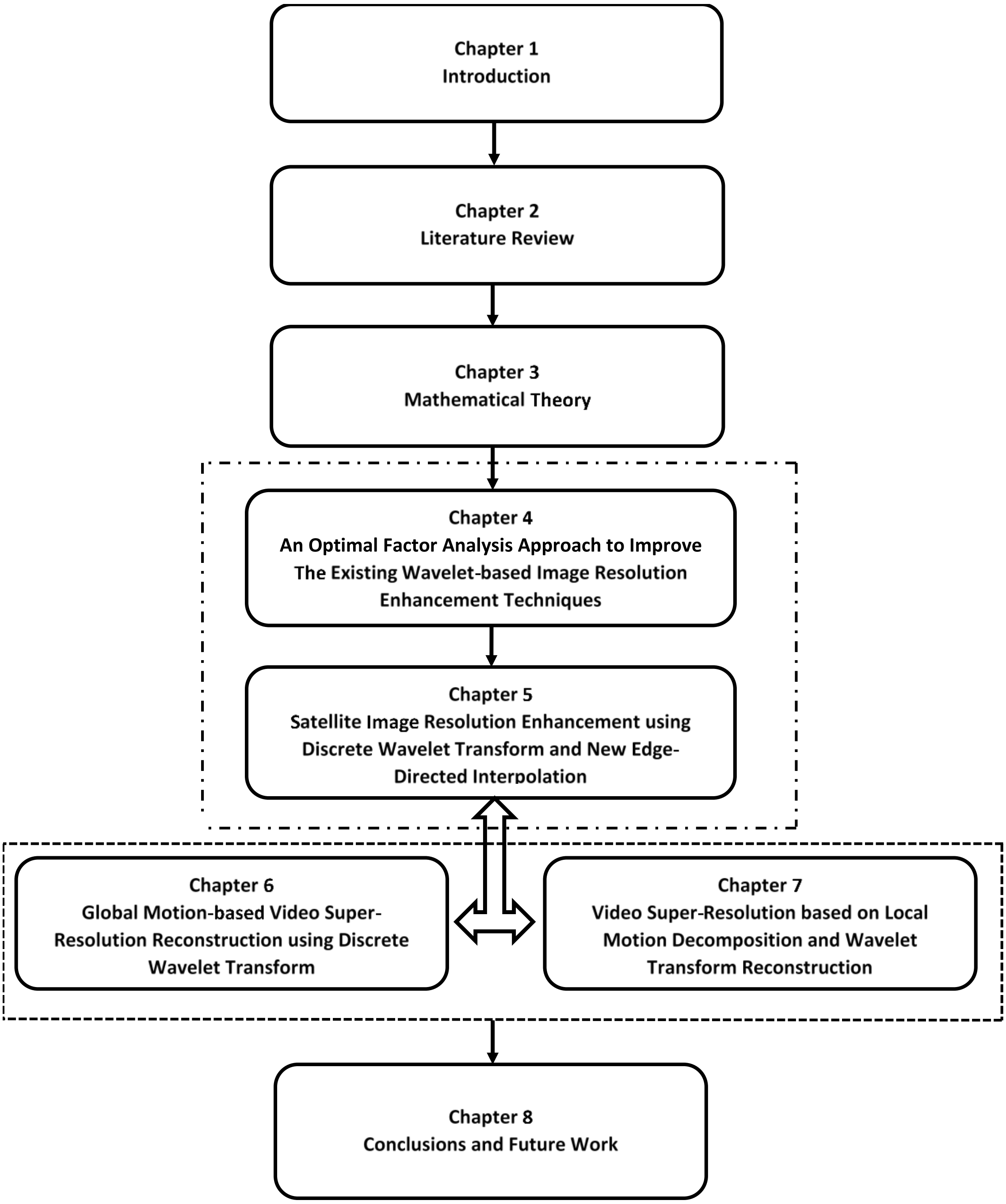


Figure 1-3 Thesis structure

Chapter 6 describes the development of a robust wavelet-based SR algorithm based on a combination of the DWT-NEDI method, developed in Chapter 5 with, spatial-domain SR reconstruction methods. The new algorithm is to address the SR reconstruction problem of the observed noisy LR video frames taken from a moving camera with global motion, in order to increase the spatial resolution, and remove the noise and aliasing artefacts. The content of Chapter 6 has been published (see Publication 3 in the Publication List)

Chapter 7 describes the designing of a new wavelet-based SR framework based on an adaptive local motion decomposition and wavelet transform reconstruction (ALMD-WTR). This is to solve the challenge of the SR problem for the real-world video sequences with complex local motion between frames. The ALMD-WTR approach is based on the DWT-NEDI method developed in Chapter 5 and the global-based SR method developed in Chapter 6.

Chapter 8 describes the accomplishment of the research aim and objects. It explains the key challenges, contributions and limitations of the research. It also highlights the overall and most significant conclusions, and suggests recommendations for future work.

2 Literature Review

2.1 Introduction

In general, most SR approaches can be classified into two main categories: reconstruction-based SR approaches and learning-based SR approaches. Fundamentally, reconstruction-based SR approaches exploit the additional information in each observed LR image provided by the sub-pixel displacements, and then reconstruct a HR image or a set of HR images. The essential idea behind learning-based SR approaches is to model the relationship between LR and HR images with the available image pairs in the database and then infer HR image from input LR images within the established model. This research focuses on SR as a reconstruction problem.

SR has been an active research topic over the last two decades. Early consideration of SR mainly followed work by Tsai and Huang [20] in 1984, based on the shift and aliasing properties of the Fourier transform. Subsequently, Park et al. [1], and Borman and Stevenson [21], [22], provided a comprehensive overview of exiting SR reconstruction techniques and described how reconstruction-based SR methods can be classified based on a variety factors. These factors include the domain utilised, the number of LR images involved, and the kind of adopted reconstruction method. In terms of the domain deployed, SR algorithms can initially be classified into three categories: the spatial-domain, the frequency-domain, and the wavelet-domain. Based on the number of LR images included, spatial-domain and wavelet-domain-based algorithms can be further divided into two classes: single image or multiple images. Based on these classifications, Figure 2-1 shows a taxonomy of the surveyed SR methods.

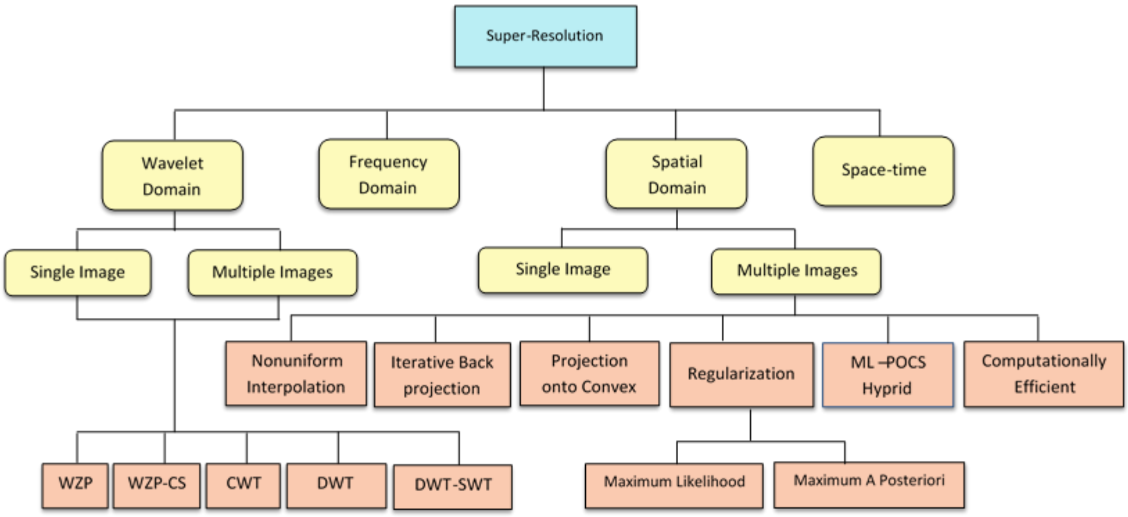


Figure 2-1 A taxonomy of the surveyed SR methods

Reconstruction-based SR approaches consist of two main phases: image registration and image reconstruction. In image registration, a sequence of LR images is aligned with sub-pixel accuracy. In image reconstruction, the registered images are combined to reconstruct a HR image [15]. Accurate sub-pixel image registration is required in order to correctly reconstruct the HR image. However, precise motion estimation is a challenging task for practical video sequences with relatively complex local motion between frames. Registration algorithms can be conducted either in the spatial-domain or the frequency-domain. Frequency-domain-based registration algorithms are usually limited to global motion models, while spatial-domain-based registration algorithms largely allow for more general motion models. After registration, most reconstruction-based SR algorithms reconstruct the high-frequency image information in either spatial-domain or frequency-domain. Although the frequency-domain-based reconstruction algorithms are usually simple and computationally inexpensive, they are limited to global translational motion and linear space-invariant blur. On the other hand, the spatial-domain-based reconstruction algorithms can accommodate both global and local motion, linear space-variant blur and noise, but they are generally computationally expensive. Typical spatial-domain-based SR approaches include: Non-uniform interpolation method; Iterative back-projection (IBP) method; Projection onto convex sets (POCS) method; Regularization-based methods, which include Maximum likelihood (ML) method and Maximum a posteriori (MAP) method; Hybrid ML/MAP-POCS method; and Computationally efficient method. Among these two domains, the wavelet-domain-based SR approach has more recently emerged as a powerful tool for many image and signal processing applications. The wavelet-domain-based approach is able to integrate merits from both the frequency-domain and spatial-domain. The wavelet-domain proves to be a natural setting for various real-world signals applications, involving resolution enhancement, estimation, detection, classification, compression, synthesis, and other applications. The wavelet transform (WT) has several attractive primary properties, such as locality, multi-resolution, and compression that make it a natural for analysing real-world signals [16]. The wavelet transform decomposes image data into different frequency sub-bands, and then studies each sub-band with a resolution matched to its scale [17]. The advantage of wavelet decomposition is that image tendency at different scales can be separated and analysed [18]. The fundamental idea behind wavelet-based SR approaches is to estimate the unknown wavelet coefficients of high-frequency sub-bands in order to reconstruct the HR image from the given LR images. There are different types of wavelet transforms, such as discrete wavelet transform (DWT), complex wavelet transform (CWT), and stationary wavelet transform (SWT). DWT is one of the recent wavelet transforms being used for resolution enhancement and many image/video applications.

2.2 Frequency-Domain-based Approaches

Tsai and Huang [20] proposed the first frequency-domain-based SR approach to reconstruct a HR image from several shifted, down-sampled and noise-free LR images. They formulated the system equations that relate the HR image to the observed LR images by estimating the relative shifts between LR images. The frequency-domain formulation is based on three basic properties: i) the shifting property of the Fourier transform (FT); ii) the aliasing relationship between the continuous Fourier transform (CFT) of an original HR image and the discrete Fourier transform (DFT) of observed LR images; and iii) an original image is assumed to be bandlimited. Kim et al. [23] extended this method to noisy and blurred LR images by proposing a weighted least squares solution to solve the set of linear equations in the frequency-domain. It was based on the assumption that the blur and noise characteristics are the same for all LR images. This method was further extended by Kim and Su [24] to the case where different blur operators for each LR image are considered by employing the Tikhonov regularization algorithm. A major advantage of the frequency-domain-based SR methods is that they are usually theoretically simple and computationally inexpensive. They are also intuitive tools to enhance the detailed information. However, these methods are inadequate when dealing with real-world applications as they are limited to global translational motion and linear space-invariant blur (LSI) during the image acquisition process.

2.3 Spatial-Domain-based Approaches

A major advantage of the spatial-domain-based SR methods is that they can tackle real-world applications better by accommodating: both global and non-global motions, linear space-variant (LSV) blur, and noise during image acquisition process. However, these methods are generally computationally expensive.

2.3.1 Nonuniform Interpolation Approach

Most of the SR reconstruction methods consist of three basic tasks: (i) registration or motion estimation, (ii) interpolation, and (iii) restoration. These tasks can be performed separately or sequentially depending on the reconstruction method employed. A nonuniform interpolation approach is one of the most intuitive methods for SR reconstruction. The three stages illustrated in Figure 2-2 are implemented sequentially in this approach. *Registration* is the process of geometrically aligning images (the reference and observed images) of the same scene acquired at different times, from different views, and/or by different sensors. Motion estimation/registration is a crucial stage in SR reconstruction [13]. The basic idea of motion estimation is that the relative displacements between the reference LR image and each of the observed LR images are estimated with sub-pixel accuracy. Accurate sub-pixel image registration is a fundamental requirement to the success of any SR reconstruction algorithm [14]. As the relative displacements are arbitrary, the composition of LR images will not coincide with a uniformly sampled HR grid. Nonuniform interpolation stage is therefore essential to produce a uniformly spaced HR image from a nonuniformly spaced composition of LR images [1]. *Interpolation* aims to estimate values at unknown locations using known data and generate HR image. It is a necessary process used widely in several applications of image processing, such as image resizing, image rotation, and sub-pixel image registration [19]. In the third stage, *restoration* is performed on the interpolated image to reduce noise and de-blurring by applying any deconvolution method [1].

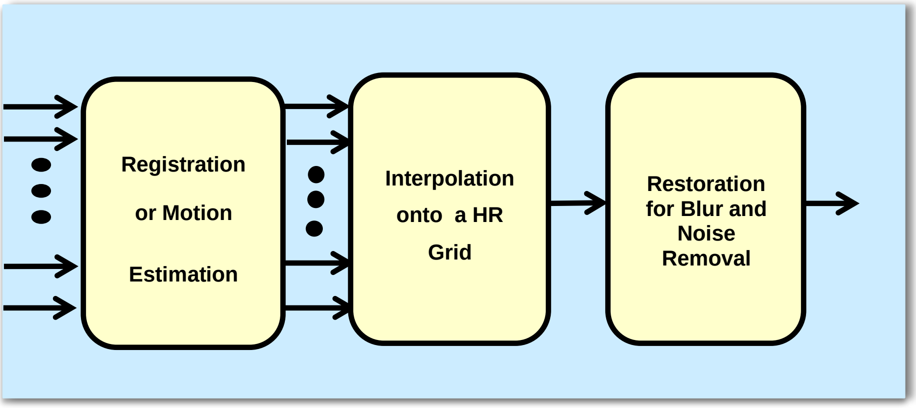


Figure 2-2 The three stages for SR reconstruction methods

Figure 2-3 is an example of global translation and shows that the interpolation process in a HR grid is needed when the sub-pixel of motion is not located on the HR grid. In this Figure, a circle is the original HR image x, and a triangle as well as a diamond are globally translated versions of x. When the down-sampling factor is 2, a diamond has $(0.5,0.5)$ sub-pixel shifts, but a triangle has a shift of less than $(0.5,0.5)$. In this case, interpolation is required only for a triangle, while it is not needed for a diamond [1].

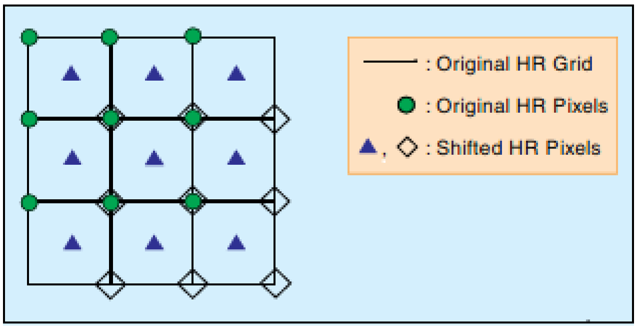


Figure 2-3 The requirement of interpolation stage in a HR grid [1]

Ur and Gross [25] implemented a nonuniform interpolation of several spatially shifted LR images based on the generalized multichannel sampling theorem of Papoulis [26] and Brown [27]. They assumed that the point-spread function (PSF) of the detectors and the relative displacements between LR images are accurately known. Alam et al. [28] developed a method for real-time infrared image registration and multi-frame SR reconstruction. They employed a gradient-based registration algorithm to estimate the random translational shifts between the acquired frames and developed a weighted nearest neighbour method to locate the frames on a uniformly spaced HR grid. The Wiener filter is used by Alam et al. [28] to reduce the effects of blurring and noise caused by the limitations of imaging systems. Nguyen and Milanfar [18] proposed an efficient wavelet-based interpolation SR reconstruction method by exploiting the interlaced structure and regularity of the sampling grid in SR. The method proposed significantly minimises the computational complexity for 2-D interlaced data to only twice that for 1-D data. Pham et al. [29] proposed a so-called structure-adaptive normalised convolution (SANC) method for fusion of irregularly sampled LR images. This method uses a normalised convolution (NC) method [30] which models the local signal through a projection onto a set of polynomial basis functions. Different from the traditional NC, a Gaussian applicability function is adapted to extend along local linear structures. This enables sufficient samples of the same intensity and gradient information to be gathered for a better analysis, which leads to improved signal-to-noise ratio (SNR) and minimises diffusion across edges. A robust signal certainty is also adapted to reduce the errors of outliers due to sensors or mis-registration. The advantage of the nonuniform interpolation method is that it has relatively low computational cost, which is suitable for real-time applications. However, degradation operators are only applicable when all LR images have the same blur and noise characteristics.

2.3.2 Iterative Back-Projection (IBP) Approach

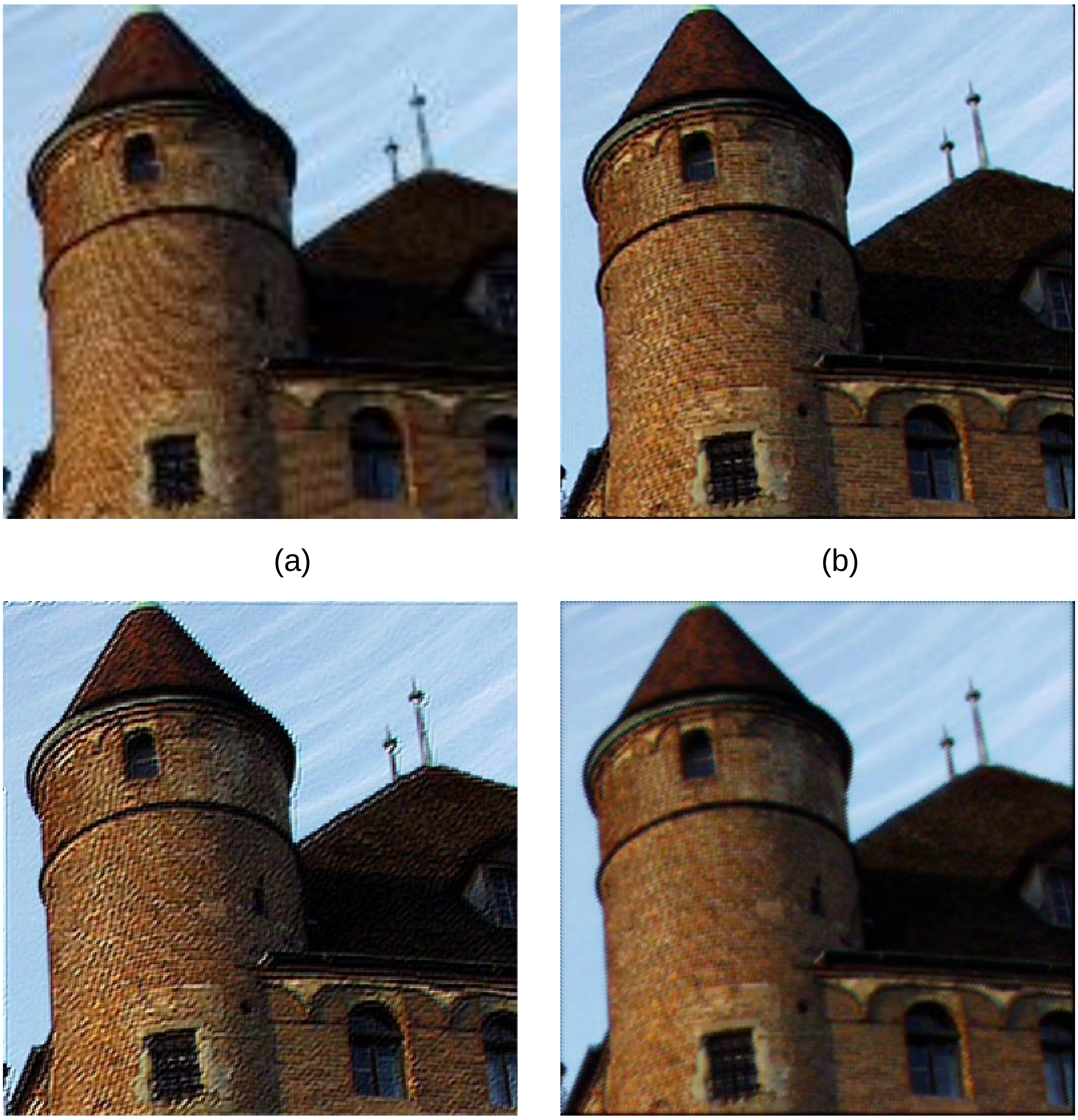
Peleg et al. [31] formulated the iterative back-projection (IBP) method by estimating an initial guess of a HR image, and simulating the imaging process to achieve simulated LR images. They improved this guess by iteratively back projecting it to the error between the simulated and the observed LR images, and updating the result according to the minimal error. The algorithm is applied to the case of a simple uniform translational motion between the observed LR images. Although this method produced good results for noise-free images, the performance for noise images is limited. Later, Irani and Peleg [32], [33] further developed the IBP method by applying a simple uniform translational and rotational motion between the LR images; but the method can also be applied for other motion models, such as perspective motion and multiple motions in the image. The proposed method resembles the IBP method employed in computer aided tomography (CAT). Irani and Peleg [34], [35] also proposed detecting and tracking multiple moving objects over several frames to improve its resolution using the IBP method. A more general motion model such as an affine model is employed to estimate the motion of objects. The advantage of the IBP method is that it can accommodate both global translational and rotational motions. However, the solution might not be unique due to the ill-posed nature of the SR reconstruction problem, and the selection of some parameters is usually difficult.

2.3.3 Projection onto Convex Sets (POCS) Approach

The projection onto convex sets (POCS) method represents an alternative iterative approach to include prior knowledge about the solution into the SR reconstruction process when the SR problem is posed in set-theoretic formulation. Set-theoretic estimation produces a solution whose property is compatible with the information obtained from the observed LR images and a priori knowledge. This information is corresponded with the constraint sets in the solution space where the intersection of these sets provides the acceptable solution [36]. Stark and Oskoui [37] formulated the POCS method, and accounted for the sensor blur due to the physical dimensions of the LR sensor.

Tekalp et al. [38] extended the POCS formulation to consider the sensor noise in addition to the sensor blur. These methods assume that the relative motion between the LR images is translations only and the aperture time of the camera is ignored. Patti et al. [39] then further developed the POCS algorithm to take into account motion blur due to nonzero aperture time, space-variant blur, sensor blur due to nonzero aperture size, sensor noise, and sampling over an arbitrary lattice. As the original algorithm was limited to the case of a single moving object in the scene, it was subsequently extended by Eren et al. [40] to the case of scene containing multiple moving objects using a validity map and/or a segmentation map. The validity map permits robust reconstruction in the case of inaccurate motion estimation, whereas the segmentation map enables an object-based resolution enhancement approach. Patti and Altunbasak [41] proposed improving the accuracy of the POCS-based SR reconstruction techniques in two ways. Firstly, they improved the discretization of the continuous image formation model by employing higher order interpolation methods. Secondly, they modified the constraint sets by including regularization constraints into the inversion process to decrease the ringing effects in the neighbourhood of edges created in the HR image estimate. The POCS method benefits from the utilisation of the powerful image formation model and an appropriate inclusion of a priori knowledge. The disadvantages on the other hand are lack of a unique solution, slow convergence rate and an expensive computational cost. The reconstruction results of spatial-domain-based SR approaches, such as a nonuniform interpolation approach using the SANC method, IBP, and POCS, are illustrated in Figure 2-4. In this simulation, four LR images with a size of 128 × 128 pixels are obtained from the Graphical User Interface (GUI) software, where the blur and noise characteristics are the same for all LR images. Figure 2-4(a) shows one of the four LR images and (b)-(d) show super-resolved images generated from the fusion of the four LR images using the SANC, IBP, and POCS reconstruction methods respectively, where the Keren method was used for registration. A significant improvement with shaper results is observed in Figure 2-4(b) when compared with (c) and (d). A moderate improvement combined with aliasing effect is observed in

Figure 2-4(c) and a moderate improvement combined with blurring effect is observed in (d).



(a) (b)

(c) (d)

Figure 2-4 An example of SR reconstruction results using different reconstruction methods (a) one of the observed LR images, (b) the HR image produced by SANC method for reconstruction and Keren method for registration, (c) the HR image produced by IBP method and Keren method, (d) and the HR image produced by POCS method and Keren method.

2.3.4 Regularized-based Approach

SR image reconstruction is generally an ill-posed inverse problem due to an adequate number of LR images, and ill-conditioned motion estimation and blurring effects. Thus, the solution to the ill-posed problem is not unique. When the solution is not unique, some additional information is needed to constraint the solution. Therefore, regularized-based approaches are proposed to regularize the inversion of this ill-posed estimation problem. Regularization methods incorporate a priori knowledge of the target HR image in order to compute an approximate solution. The target HR image can be estimated by some statistics of its probability distribution. In this section, the stochastic regularization approaches are presented, including a Maximum likelihood (ML) estimation and Maximum a posteriori (MAP) estimation.

2.3.4.1 Maximum Likelihood (ML) Approach

Tom and Katsaggelos [42] proposed a Maximum likelihood (ML) SR image estimation algorithm to simultaneously estimate the sub-pixel displacements, the noise variances of each image, the HR image, and the power spectra of the HR image. They employed the expectation maximization (EM) algorithm to solve the ML estimation. Capel and Zisserman [43] used ML and MAP estimators for the SR enhancement of video mosaics. The transformation estimation algorithm was based on a general homography model. ML estimation is a special case of MAP estimation without prior term. Because the inclusion of a priori information is necessary for the solution of ill-posed inverse problem, MAP estimation is preferable to ML.

2.3.4.2 Maximum A Posteriori (MAP) Approach

Stochastic regularized approaches (Bayesian estimation methods) solve the ill-posed SR inverse problem by regarding SR reconstruction as a statistical estimation problem. They provide a powerful and appropriate way to include a priori knowledge essential for the successful solution of the ill-posed SR inverse problem. A Bayesian approach is used when the posteriori probability density function (PDF) of the HR image can be modelled. The HR image x and the observed LR images y_k can be treated as stochastic variables.

The MAP estimation method is one of the probabilistic SR reconstruction formulations and solutions to incorporate various regularizing constraints into the inversion process [1], [2]. Schultz and Stevenson [44] introduced a Bayesian estimation method for a single frame image expansion that keeps the edges and other discontinuities of the original image. They employed a MAP method to compute the expanded image and addressed two problems: a constrained optimization problem for expanding the image that is a noise-free; and an unconstrained optimization problem for expanding image data containing white Gaussian noise. The same authors [45] then extended their work and developed the MAP method for SR reconstruction from a LR video sequence. They utilised a discontinuity-preserving Huber-Markov random field prior model within the MAP formulation. They also employed a hierarchical block matching registration technique to estimate the sub-pixel motion parameters between frames. Hardie et al. [46] introduced a joint MAP framework to estimate simultaneously the image registration parameters and the HR image. They iteratively updated the registration parameter estimates with respect to the HR image using a cyclic coordinate-descent optimization process. Shen et al. [47] used a joint MAP formulation to integrate motion estimation, segmentation and super-resolution together for the more complex multiple moving objects problem. They also applied a cyclic coordinate-descent process to solve the formulation, in which the motion fields, the segmentation field, and the HR image are considered as unknown and estimated jointly by using the available data. Li et al. [48] proposed applying two new regularization items, namely locally adaptive bilateral total variation and gradient consistency, to preserve both edges and flat regions of the LR images. Capel and Zisserman [49] proposed a MAP estimator based on a Huber-edge penalty function, and an estimator regularized based on the total variation norm for the SR enhancement of text images. They proved that the ML estimator is extremely sensitive to small amounts of errors, such as noise, registration errors, and PSF errors. In general, the ML method only considers the relationship among the observed LR images and the original HR image without prior information while the MAP method considers both.

2.3.5 (ML/MAP-POCS) Hybrid Approach

Elad and Feuer [50] proposed a general hybrid approach which combines the advantages of the regularization methods (ML or MAP) and the POCS method to solve the SR reconstruction problem. The hybrid algorithm effectively combines all a priori information and guarantees the single optimal solution.

2.3.6 Computationally Efficient Approach

SR is a computationally expensive problem and involves tens of thousands of unknown pixel values. For instance, super-resolving a sequence of LR images with a size of 128×128 pixel values by a scale factor of 4 in each spatial dimension requires 512×512 = 262,144 unknown pixels in the HR estimate. Moreover, the matrix system is ill-conditioned and typically underdetermined problem which can trigger system noise and blurring artefacts. Therefore, to reduce the computational and numerical issues, it is important to develop efficient SR algorithms that deal with real-time applications. Nguyen et al. [51] presented a block circulant preconditioner to accelerate the conjugate gradient (CG) method for solving the Tikhonov-regularized SR problem. They also utilised the generalized cross-validation (GCV) method to automatically compute the regularization parameters and extended the derivation of the GCV to the underdetermined problems. Elad and Hel-Or [52] proposed a fast and effective SR reconstruction algorithm that separates the de-blurring from the fusion stage. To reduce the computational cost, they dealt with a special SR case where the geometric warps between the measured images consist of pure translations; the blur is space-invariant and the same for all the measurements, and the additive noise is white. Although the fusion process is achieved by a simple noniterative algorithm, it can preserve the optimality of the SR reconstruction process in the ML sense. Elad and Hel-Or also provided a proper mathematical justification of the noniterative method for this simple case of the additive Gaussian noise. Although these assumptions limit the utility, they are very practical in some cases, such as the case in a video sequence where the scene is stationary and the images are acquired with slight translations. Zomet et al. [53] introduced a successful SR algorithm which is robust to outliers due to local model inaccuracies, such as moving objects and highlights, at a low computational load. They used a robust median-based estimator to estimate a summation in an iterative process and discard inconsistent measurements. However, the limitation of this method is that it lacks an adequate mathematical justification. A robust SR method was presented by Farsiu et al. [54] based on the use of L1 norm minimisation both for the regularization and the measurement terms. They proved that the additive Gaussian noise model is not satisfactory for SR problem, then proposed an appropriate mathematical justification for the noniterative image fusion algorithm based on shift and add method and L1 norm for more general noise models. Farsiu et al. [55] also proposed a fast and robust SR algorithm using the L1 norm and robust regularization based on bilateral variation prior to deal with different data and noise models. This method is robust to motion errors, blur estimation, and outliers, resulting in images with sharp edges.

2.4 Wavelet-Domain-based Approaches

The wavelet transform (WT) is an effective tool that decomposes data or functions or operations into different frequency components, and then studies each component with a resolution corresponded to its scale [17]. The advantage of wavelet decomposition of a signal is that signal characteristics at different scales can be separated and examined. Global characteristics can be examined at coarser scales, while local characteristics can be analysed at fine scales [18]. The wavelet transform has advantages over the Fourier transform in analysing real-wold signals where the signal contains discontinuities [12]. Wavelet transforms depend on small waves, known as wavelets, produced from a single mother wavelet function using the property of dilations and translations. Wavelet transforms and wavelets are a relatively recent imaging tools being applied rapidly in a wide range of image processing applications, including enhancement, restoration, de-noising, registration, and segmentation [19]. Recently, there has been much development on wavelet-based SR reconstruction and resolution enhancement methods. Estimating the unknown wavelet coefficients in sub-bands containing high-pass frequency spatial information is the essential target of wavelet-based algorithms to reconstruct the HR image from the given LR image/images. A simple approach, called wavelet zero padding (WZP), is to recover an approximation to the HR image by filling the unknown wavelet coefficients of high-frequency sub-bands with zeros and then applying the inverse wavelet transform (IWT). Although this method is able to surpass the conventional interpolation methods, it commonly introduces artefacts, such as smoothing and ringing, into the reconstructed HR image. To overcome this limitation, many advanced methods have been introduced to estimate the wavelet coefficients of high-frequency sub-bands. The existing literature on wavelet-based resolution enhancement and SR reconstruction approaches describes different types of wavelet transforms in both the single frame (image) case and multi-frame (video) case, which are introduced below:

2.4.1 Single Frame (Image) Resolution Enhancement Approaches

2.4.1.1 Extrema Evolution Approach

Chang et al. [56] and Carey et al. [57], estimated only the coefficients of high-frequency sub-bands with large magnitudes as the evolution of the wavelet coefficients among the scales while it is difficult to estimate the other small coefficients. The performance is fundamentally affected by the signs of the estimated coefficients being copied directly from the signs of the parent coefficients without any attempt to estimate the actual signs. However, the accepted fact is that there is very low correlation between the signs of the parent coefficients and the estimated coefficients. Therefore, the signs of the estimated coefficients using extrema evolution methods cannot be relied upon.

2.4.1.2 Hidden Markov Tree (HMT) Approach

The hidden Markov tree (HMT) approach in the wavelet-domain is able to precisely model the statistical structure of real-wold signals by exploiting the statistical relationships between wavelet coefficients at different scales. Crouse et al. [16] introduced a statistical approach for signal processing by using wavelet-domain hidden Markov models (HMMs). The HMM framework is capable of matching the non-Gaussian statistics for individual coefficients and characterising the statistical dependencies between coefficients. Subsequently, Kinebuchi et al. [58] proposed a hidden Markov tree-based method using HMM approach for image enlargement. Zhao et al. [59] presented an extended version of this method. All these methods are based on modelling the unknown detail coefficients by the mixed Gaussian distributions (states), which are symmetrical around zero. The motivation comes from the fact that the coefficient distributions could be modelled by a Gaussian mixture model, as they have high density at zero and are heavy tailed. HMT-based methods are used to determine the most probable state for the estimated coefficient. However, their main drawback is that the performance is affected by the sign changes between the scales not being taken into account. To reduce the drawback of HMT-based methods which affect the accuracy of coefficient sign estimation and consequently the resulting image quality, a refined HMT-based image resolution enhancement method was proposed by Temizel [60]. Due to the fact that the coefficient sign and magnitude information which are statistically independent could be separated, the coefficient magnitude estimation is separated from the sign estimation. The magnitude parameters are estimated using HMT methods, and the sign estimation is modelled based on the work of Temizel and Vlachos [61]; this demonstrated that there is higher correlation among the coefficients between high-pass filtered wavelet LR image and the high-frequency sub-bands.

2.4.1.3 Cycle-Spinning (CS) Approach

Temizel and Vlachos [62] developed a nondirectional cycle-spinning (CS) technique, termed WZP-CS, as an effective method towards reducing ringing artefacts. It was based on obtaining an initial estimate of the unknown HR image using WZP, and generating a number of LR images from this initial estimated HR image using spatial horizontal and vertical shifts, wavelet transform, and discarding the high-frequency sub-bands. This is followed by applying WZP to all those LR images and averaging these intermediate HR images to give the final reconstructed HR image. However, ringing artefacts not only occur in the neighbourhood of edges, they are particularly correlated with the orientation of edges. A directional CS technique, which can refine better edge orientation and prevent ringing artefacts, was introduced in [63] by

Vlachos. Additionally, it can reduce the computational complexity compared with a nondirectional CS technique. A further improvement of this method could be obtained by applying a CS and edge rectification technique [64] proposed by Temizel and Vlachos.

2.4.1.4 Complex Wavelet Transform (CWT) Approach

A one-level complex wavelet transform (CWT) of an input image results in two complex-valued low-frequency sub-bands and six complex-valued high-frequency sub-bands. The high-frequency sub-bands are constructed by direction-selective filters with high magnitude responses in the presence of image features oriented at +75°, +45°, +15°, -15°, -45°, and -75°. The advantages of CWT is that it has shift-invariant property and good directional selectivity. Reeves and Kingsbury [65] introduced a dual-tree complex wavelet transform (DT-CWT). The DT-CWT is a combination of two real-valued decimated DWT. Recently, Demirel and Anbarjafari [66] proposed a DT-CWT technique for resolution enhancement of satellite images. One-level DT-CWT decomposes an input LR image into different frequency sub-bands, and then the high-frequency sub-bands are interpolated using bicubic interpolation. In parallel, two magnified real-valued sub-bands are produced by interpolation and a shifted version of the input LR image in the horizontal and vertical directions; these are used as the real and imaginary components of the interpolated complex low-frequency sub-band. Finally, a super-resolved image is generated by combining all these interpolated sub-bands through the inverse DT-CWT. More recently, Iqbal et al. [67] proposed a DT-CWT technique based on a non-local means filter and Lanczos interpolation in order to improve the performance. Although DT-CWT is approximately shift-invariant and has improved directional resolution, it generates artefacts in the interpolated high-frequency sub-bands. Therefore, non-local means filtering is used to reduce these artefacts. Jagadeesh and Pragatheeswaran [68] used an edge-directed interpolation (EDI) method as an alternative interpolation method of high-frequency sub-bands obtained by DT-CWT to improve the performance.

2.4.1.5 Discrete Wavelet Transform (DWT) Approach

The discrete wavelet transform (DWT) has emerged as a dominant and powerful new tool for resolution enhancement and is being employed in many image and signal processing applications. The advantage of DWT lies in isolating and recovering the high-frequency information of an image that is lost due to the limitations of imaging systems [19]. The DWT provides an adequate information for local analysis and synthesis of an image [12]. However, the decimated DWT is shift-variant and, thus, suppression of wavelet coefficients because of the decimation operation exploited in the transform introduces artefacts into the image which emerge as ringing in the neighbourhood of discontinuities. The fundamental concept of DWT is the decomposition of a given image into four frequency sub-bands: low-low (LL), low-high (LH), high-low (HL) and high-high (HH); achieved by using dilations and translations by a single wavelet function, referred to as the mother wavelet. Acharya et al. [69] proposed using one-level DWT to separate an input LR image into four frequency sub-bands, and then the LL sub-band is replaced by the input image, the high-frequency sub-bands LH and HL are up-sampled by inserting zeros between successive rows and columns, and HH is discarded. A high-resolution image is generated by performing the inverse DWT on these approximated sub-bands. Recently, a Demirel-Anbarjafari Super-Resolution (DASR) technique was proposed [70], in which the decomposed three high-frequency sub-bands and the input image are interpolated using the bicubic interpolation. Upgrading from the DASR, Demirel and Anbarjafari [71] introduced a DWT-Diff method for resolution enhancement of satellite images in which the estimated high-frequency sub-bands are further enhanced by considering the difference between the input LR image and the interpolated LL sub-band, and then adding the difference image into the estimated high-frequency sub-bands.

2.4.1.6 Discrete Wavelet Transform (DWT) and Stationary Wavelet Transform (SWT) Approach

Demirel and Anbarjafari proposed a DWT-stationary wavelet transform (SWT) technique based on DWT and SWT [72], which introduced an intermediate process by adding the high-frequency sub-bands obtained through SWT of the input image with the high-frequency sub-bands obtained through DWT to improve the performance. SWT is similar to DWT but it does not cause down-sampling and, as a result, each sub-band has the same size as the input image. A 2-D and 3-D technique was proposed [73] for resolution enhancement of medical images based on DWT and SWT. The 2-D version of this technique outperforms the results in [72] by computing the mean of the corrected high-frequency sub-bands, and using a shape function to enhance the edges of the input LR image. In the 3-D version, 3-D DWT is employed to decompose a 3-D LR image into eight different sub-bands and then 3-D bicubic interpolation is used to up-sample the high-frequency sub-bands. Additionally, 3-D SWT is applied to these interpolated coefficients and their mean are considered to correct the estimated coefficients. Finally, the 3-D IDWT is applied to create a shaper enhanced HR image.

2.4.2 Multi-Frame (Video) Super-Resolution Approaches

For the multi-frame case, Izadpanahi and Demirel [74] introduced a resolution enhancement method based on DT-CWT and EDI for LR videos to improve the performance. Izadpanahi and Ozcinar [75] presented a SR technique using DWT and bicubic interpolation of LR video frames. They applied an illumination enhancement method based on singular value decomposition before the registration process of the LR frames to reduce the illumination inconsistences between the reference frame and each of the other LR frames. One-level DWT decomposes each reference frame of the input video frames into four frequency sub-bands and the bicubic interpolation is applied to the high-frequency sub-bands of the consecutive frames. On the other hand, the Irani and Peleg registration method is performed on the LR frames to obtain the LL sub-band of DWT. Finally, all these sub-bands are composed using IDWT to produce a super-resolved frame. Anbarjafari et al. [76] proposed a SR technique for LR video sequences using DWT and SWT. For proper registration, they also performed the illumination enhancement method by using the highest singular value to decrease the illumination differences between the frames. One-level DWT and SWT decompose each reference frame of the video sequences into four frequency sub-bands, and then the high-frequency sub-bands obtained by DWT are enlarged using a SR reconstruction method. Furthermore, those high-frequency sub-bands are enhanced by the sub-bands obtained through SWT, as these sub-bands contain more information. In parallel, the LR frames are registered using the Vandewalle method to produce the LL sub-band of DWT.

For motion-based video SR, Izadpanahi and Demirel [77] proposed a SR technique using DT-CWT and new edge directional interpolation (NEDI) to improve the resolution of LR video sequences. They divided each frame into stationary and motion regions for better registration. A SR reconstruction technique is performed on the stationary regions and the motion blocks extracted by the motion detection algorithm to obtain the LL sub-bands of DT-CWT. On other hand, DT-CWT process followed by NEDI method is used to generate the interpolated high-frequency sub-bands. Izadpanahi and Demirel [78] also presented a SR technique based on detecting the motion and static regions of the LR frames. They divided the static and motion regions into small blocks using an optical flow motion estimation algorithm between a reference frame and its neighboring frames for accurate registration. Additionally, an adaptive threshold is performed on the motion blocks in order to separate the occluded blocks with inaccurate motion from non-occluded blocks with accurate motion for more accurate local registration.

2.5 Interpolation Methods

Interpolation is one of the most commonly used techniques for image resolution enhancement. Fundamentally, *image interpolation* is a process of estimating values at unknown locations using known data [19]. Interpolation is a basic process used extensively in several digital image processing applications such as image rotation, sub-pixel image registration, and image geometric corrections. There are four well-known conventional interpolation methods: nearest neighbour, bilinear, cubic spline, and Lanczos [79]–[84]. Nearest neighbour is the simplest interpolation method, where the intensity value of the new location is assigned as that of the old location which is the nearest neighbour to the new point. Although this method is simple to implement, it results in annoying artefacts, such as distortion of straight edges. A more frequently used method is bilinear interpolation, where the four nearest neighbour points are used to estimate the value at a new location by taking a weighted average of these points. This method produces much better results than nearest neighbour with a moderate increase in computational load [81], [82]. Cubic spline interpolation methods [79]-[83] have been researched extensively and received much attention. Hou and Andrews [79] investigated the cubic B-spline interpolation function, which sometimes is also known as cubic spline, as a tool in image interpolation to smooth the interpolated image. Keys [80] adopted the cubic convolution interpolation, which is also referred to as bicubic interpolation and HR cubic spline interpolation in [81], [82] and [83]. Bicubic interpolation is a more complex method, in which sixteen nearest neighbour points are used to estimate the intensity value of the new point by taking a weighted average of these points. This method is more efficient and accurate than nearest neighbour, bilinear interpolation and cubic-B spline interpolation, as well as giving slightly sharper results in terms of preserving fine details in the original image [80]. It is generally the most preferred in commercial image editing programs [19]. Lanczos interpolation also produces good results because of its capability to detect edges and linear features. It delivers a good compromise in terms of reduction of aliasing, sharpness of edges, and suppression of ringing artefacts [67] and [84]. However, these linear methods cannot handle the fast-growing statistics around edges and accordingly yield interpolated images with blurred edges and undesirable artefacts. To address this problem, other nonlinear interpolation methods have been developed to improve the visual quality of the reconstructed images over linear interpolation methods by taking edge information into account; these techniques include edge-directed interpolation (EDI) [85], [86], and new edge-directed interpolation (NEDI) [87]. Allebach and Wong [86] developed an EDI method [85] through emphasising the visual integrity of the edges. The EDI method consists of two stages: rendering and correction. The rendering stage, termed edge-directed, is based on modifying the interpolation procedure to match a source model and prevent smoothing across edges. The correction stage is based on correcting the mesh values on which the interpolation is based. Li and Orchard [87] proposed a NEDI method, known as also an orientation-adaptive interpolation, to estimate local covariance coefficients from a LR image and then employed the covariance estimates to modify the interpolation scheme by using the geometric duality between the LR covariance and the HR covariance. The advantage of the NEDI method is that it significantly improves the subjective quality of the pixels around edges. Nevertheless, the improvements by NEDI are limited at the textures and nonlinear edges of the interpolated images. However, a lot of research has achieved enhanced performance of NEDI since the NEDI method uses a relatively simple model and hence has low computational complexity. Tam et al. [88] presented a modified NEDI method by considering a modified training window structure to eliminate the prediction error accumulation and extending the covariance values into multiple directions to mitigate the covariance mismatch problem.

2.6 Registration Methods

Image registration aims at overlying (two or more) images of the same scene of interest acquired at different times, from different viewpoints, and/or by different sensors. The extant disparities between the images are introduced because of different conditions during image acquisition. Image registration is a necessary task used extensively in various applications in remote sensing, medical imaging, and computer vision. In general, these applications can be classified into four basic groups according to the manner of image acquisition: different viewpoints (multi-view analysis), different times (multi-temporal analysis), different sensors (multi-modal analysis), and scene to model registration. A comprehensive overview of image registration methods has been presented by Zitova and Flusser [13]. Registration methods can be operated either in the spatial-domain or the frequency-domain of images. Frequency-domain registration methods are usually limited to global motion models due to the nature of the Fourier domain. Reddy and Chatterji [89] described an extension of the phase correlation method for registering images subject to translational, rotational, and scale movement. They applied a planar motion model in order to determine the shift and rotation parameters. Marcel et al. [90] also used a planar motion model to estimate the shift and rotation between the LR images. Lucchese and Cortellazzo [91] developed a frequency-domain registration algorithm to estimate planar roto-translations between two images. A planar rotation is estimated by using the property that the difference between the magnitude of the Fourier transform FT of one image, and the mirrored version of the magnitude of the FT of the other image, has a pair of orthogonal zero crossing lines. These two lines are rotated relatively to the frequency axes with an angle that is equal to half the rotation angle. The planar shifts are estimated using a phase correlation technique. Vandewalle et al. [14] presented an image registration algorithm to accurately register a series of aliased images based on their low-frequencies, thereby aliasing its free-part. They also used a planar motion model, particularly for the scenario when a set of images are captured in a short period of time with a small camera motion. It has been seen that Vandewalle's method performs better than the other frequency-domain registration methods, such as Marcel method and Lucchese and Cortellazzo method.

Spatial-domain registration methods generally allow more general motion models. They are based on the whole image or correspondence between a set of selected features. Earliest methods in the spatial-domain used correlation methods and Gaussian pyramids/multi-scale approaches. Pratt [92] developed an extension of the basic correlation measure for registering images with translational shifts. This method is based on a linear spatial pre-processing operation of one of the images before applying the correlation measure. Bergen et al. [93] developed a hierarchical framework for estimating the motion in a multi-resolution approach. Four motion models: affine flow, planar surface flow, rigid body motion, and general optical flow can be combined within this motion estimation algorithm. Keren et al. [94] developed an iterative planar motion estimation algorithm that uses different, down-sampled versions of the images in order to estimate the shift and rotation parameters based on Taylor series expansions. The goal of this pyramidal scheme is to increase the accuracy for estimation of large motion parameters. Irani et al. [95] proposed a method to detect and track multiple occluding and transparent moving objects in an image sequence using segmentation and temporal integration. The motions are computed based on parametric motion models, such as affine and projective transformations. Gluchman [96] introduced a method for registering images that are significantly rotated and translated relative to each other. A planar rotation is first determined from the gradient field distribution of the images to be registered. Then planar shifts are computed after disregarding the rotation using a phase correlation technique. Capel and Zissermann [97] established corresponding features by first determining putative correspondences using the normalised correlation algorithm and then refining them using the RANCAC algorithm [98]. Baboulaz and Dragott [15] proposed two methods for feature extraction in LR images in order to increase the registration accuracy. These methods are used to recover perfectly global features, such as image moments, as well as local features, such as step edges in LR images.

2.7 Motion Detection Methods

Motion detection is one of the greatest problem areas in human motion analysis; it not only provides segmentation of the video sequences into moving foreground objects and background for the extraction of desired moving objects, but is also a critical pre-process for many computer vision applications, including human-machine interaction, video surveillance, and object-based video encoding [99]. The existing motion detection approaches can be classified into three classes: background subtraction, temporal differences, and optical flow [100]. Background subtraction is a popular approach for motion detection, because it is relatively simple to implement and produces reasonable motion segmentation particularly for situation with a relatively static background. However, it is extremely sensitive to changes in a dynamic scene caused by illumination and extraneous events [101], [102]. Oral and Deniz [103] used the simple background subtraction (SBS) method which can segment moving objects in a video sequence by taking an absolute difference between each pixel of the incoming frame and the background reference frame, and thresholding the output to get the binary-moving objects detection mask.

However, noise tolerance, lighting changes, weather changes, and clutter motion make motion segmentation a critical process. Therefore, there is a need for a good background models that effectively update the background reference frame through generation of adaptation models. An updated background model using running average (RA) method based on a simple adaptive filter was adopted by Wren et al. [104] for compensation to the temporal changes of illumination and weather in the video sequence. Manzanera and Richefeu [105] detect the moving objects captured by a fixed camera using the sigma-delta filter. However, the optical flow method shows good approximation of the complex background, but it usually requires expensive computational complexity. For example, Liu and Sun [106] used optical flow techniques to register multiple images with sub-pixel accuracy whereas Liao et al. [107] used an ensemble of optical flow models to reconstruct the original HR frames with rich high-frequency details.

2.8 Spatio-Temporal-based SR Approaches

In the recent years, producing high space-time resolution of video sequence has become a new challenge for SR research and has thus received increasing attention. Temporal resolution is defined by the number of frames detected per second and is usually referred to as the frame-rate. It aims to recover rapid dynamic events that occur faster than video frame-rate and therefore they are not seen or captured correctly in the recorded video sequence. The temporal resolution of the camera is limited by two factors: the frame-rate and the exposure-time that determine the maximal speed of dynamic events that can be detected in video recording. There are two visual effects in video recordings which are caused by very fast dynamic events. One effect (motion blur) is due to the long exposure time of the camera, and the other effect (motion aliasing) is caused by the insufficient frame-rate of the camera. An increase of frame-rate can result in less, or even completely avoid, visual effects caused by very fast-moving objects. Traditional methods used single video to enhance the temporal resolution and exploited the complicated temporal interpolations [108]. Shechtmane et al. [109], [110], [111] extended the concept of SR to the space-time domain to increase both the spatial and the temporal resolution by combining information from multiple LR video sequences of the same dynamic scene captured by different cameras at (sub-pixel) spatial and (sub-frame) temporal displacements. This approach introduces visual trade-offs in time and space which are unique to spatio-temporal SR and results in new visual capabilities of very fast dynamic events. This approach also produces smooth up-scaling of local pixels, but cannot handle motion aliasing. Shahar [112] showed that SR in time and space can be obtained from a single natural video sequence of a dynamic scene based on recurrence of small space-time patches inside the same video at multiple spatio-temporal scales by combining information from multiple space-time patches of the same dynamic scene obtained at sub-frame accuracy. Shimano [113] introduced a temporal SR method for increasing the temporal resolution of a video from a single input image sequence by exploiting the self-similarity which means self-similar motion blur appearances in the spatio-temporal domain of videos at different temporal resolutions. Such self-similarity resolves the unconstrained problem of temporal SR by using a MAP estimation that combines both a prior probability from self-similar appearances and reconstruction constraints.

Probabilistic motion estimation algorithms [114], [115] and steering kernel regression algorithms [116], [117] have been proposed to circumvent the problem of space-time video SR by avoiding the explicit need for precise sub-pixel motion estimation. Protter et al. [114] developed a non-local means (NLM) de-noising method by measuring the similarity of image patches across space-time and giving relatively higher weights to more comparable patches. This work resulted in "fuzzy" or probabilistic estimates of motion. The main advantage of fuzzy motion estimation is to handle sequences with an arbitrary motion pattern, thus avoiding explicit sub-pixel motion estimation. This method was then extended by Cheng et al. [115] based on dividing each frame into simple areas and complex areas to improve the performance of the NLM algorithm. Takeda et al. [116] introduced a method based on the extension of the steering kernel regression (SKR) framework [117] to 3-D signals for performing video de-noising, spatio-temporal up-scaling, and SR, without the need for explicit sub-pixel accuracy motion estimation. In this method, each pixel in the sequence is estimated by a 3-D local Taylor series, using the local behaviour of its spatio-temporal neighbourhood, and the coefficients are computed by solving a local weighted least-squares problem.

2.9 Learning-based SR Approaches

Recently, learning-based SR approaches have emerged to further boost the efficiency of SR. The fundamental idea of these methods is to capture the co-occurrence prior between LR and HR patches. Machine learning algorithms consist of two main parts: learning and recovering. In the learning part, a dictionary which contains a large number of LR and HR patch pairs is constructed. In the recovering part, the LR frame is divided into overlapped patches, and each patch searches its more similar LR patch from the dictionary. The HR frame is obtained by incorporating the corresponding HR patch into the LR frame. For example, Timofte et al. [118] introduced a simple function method and the anchored neighbour regression (ANR) method. The same authors further improved the ANR method and proposed the adjusted ANR (A^+) method for fast SR [119]. These methods divide the training data into a small number of groups and learn a regression model for each group. Deep learning-based SR approaches have also appeared to improve SR results. Huang and Siu [120] proposed using a decision tree method and a SR hierarchical decision trees (SRHDT) method for improving SR performance. To better model complex image contents and details, deep learning networks methods have been introduced. Dong et al. [121] proposed a SR convolution neural network (SRCNN) method to perform a sparse reconstruction. However, this method does not exploit natural image priors and suffers from losing sharp edges. Following the SRCNN method, a deep edge guided recurrent residual method [122] was proposed to provide high-quality image SR and recover the edges by recurrent residual learning. Takeda et al. [116] introduced a method based on the extension of steering kernel regression framework to 3-D signals for performing video de-noising, spatio-temporal up-scaling and SR, without the need for explicit sub-pixel accuracy motion estimation. To generate better results, multi-dimensional kernel regression was applied. Yang et al. [123] proposed a sparse-coding method where the LR and HR patch pairs share the same sparse representation in terms of coupled dictionaries jointly trained. The sparse representation of a LR patch can be incorporated into the HR dictionary to obtain HR patch. Kang et al. [124] proposed a joint SR and de-blocking method for a highly compressed image; they learned respectively image sparse representations for modelling the relationship between LR and HR image patches in terms of the learned dictionaries for image patches with and without blocking artefacts. Zhang et al. [125] employed clustering and collaborative representation to learn the mapping functions (i.e. projection matrices) from the LR feature spaces to their HR feature spaces for fast SR. Jiang et al. [126] also addressed the problem of learning the projection matrices by introducing the non-local self-similarity and local geometry priors of the training data. However, learning-based methods usually requires a large amount of training data.

2.10 Summary and Comparison

A general comparison of advantages and disadvantages of spatial and wavelet-domains SR reconstruction approaches is presented in Table 2-1.

Table 2-1 Summary of advantages and disadvantages of spatial and wavelet-domains SR reconstruction methods.

Methods	Advantages	Disadvantages
Nonuniform Interpolation	❖ Low computational cost.	❖ Degradation operators are only applicable when all LR images have the same blur and noise.
Iterative Back-Projection (IBP)	❖ Accommodates both global translational and rotational motions.	❖ The solution might not be unique. ❖ The selection of some parameters is difficult.
Projection onto Convex Sets (POCS)	❖ Utilizes a powerful observation model. ❖ Utilizes an appropriate	❖ Lack of a unique solution. ❖ Slow convergence rate. ❖ Expensive computational

	inclusion of a priori knowledge.	cost.
Maximum Likelihood (ML) Estimation	❖ Considers only the relationship among the observed LR images and the original HR image without prior information.	❖ Without prior information.
Maximum A Posteriori (MAP) Estimation	❖ Considers the relationship among the observed LR images and the original HR image with prior information.	
(ML/MAP-POCS) Hybrid	❖ Combines effectively all a priori information. ❖ Guarantees a single optimal solution.	
Computationally Efficient	❖ Reduces the computational load. ❖ Reduces numerical difficulty.	
Wavelet Zero Padding (WZP)	❖ Simple to implement.	❖ Introduces smoothing and ringing artefacts.
Extrema Evolution	❖ Estimates only the wavelet coefficients with large magnitudes.	❖ Difficult to estimate the coefficients with small magnitudes. ❖ Signs of the estimated coefficients cannot be relied.
Hidden Markov Tree (HMT)	❖ Estimate the wavelet coefficients by the mixed Gaussian distributions.	❖ Sign changes between the scales are not being taken into account.

Cycle-Spinning (CS)	❖ Reduces ringing artifacts.	
Discrete Wavelet Transform (DWT)	❖ Isolates and preserves the high-frequency components. ❖ Easy to implement.	❖ Shift variant.
Stationary Wavelet Transform (SWT)	❖ Isolates and preserves the high-frequency components. ❖ Easy to implement.	❖ Shift variant.
Dual Tree Complex Wavelet Transform (DT-CWT)	❖ Isolates and preserves the high-frequency components. ❖ Approximately shift invariant.	❖ Difficult to implement.

A general comparison of spatial and frequency-domains SR reconstruction approaches is presented in Table 2-2.

Table 2-2 Frequency-domain SR reconstruction approaches against spatial-domain SR reconstruction approaches.

Properties	Frequency-domain	Spatial-Domain
Observation model	Frequency-domain	Spatial-Domain
Motion models	Global translation	Almost unlimited
Degradation model	Limited, LSI	LSI or LSV
Noise model	Limited	Very Flexible
Computational requirement	Low	High
A priori Knowledge	Limited	Almost unlimited
Regularization	Limited	Excellent
Applicability	Limited	Wide
Applicability performance	Good	Good

2.11 Research Gap

Based on the literature review, the following research gap were identified:

- There was a clear research gap of the benefits to integrate the information from both the spatial-domain and frequency-domain for solving the SR reconstruction problem. The wavelet-based approach covers both of them together to bring in advantages to the SR reconstruction problem and boost the resolution limit of LR observations.
- There was a clear research gap to identify the SR reconstruction problem in the wavelet-domain in comparison with the spatial and frequency-domains in terms of the associated mechanisms, involved limitations, and achieved challenges.
- There was a research gap to clarify the difference between the wavelet-based SR reconstruction and wavelet-based resolution enhancement in terms of the associated implementation, the parameters that influence on the performance, and the quality of produced HR image.
- Although the SR reconstruction is based on reasonable assumptions or prior knowledge about the observation that maps the original HR image to the observed LR images, there was a research gap of the mechanism to generate the observed LR images. There was no discussion on the mechanism to produce LR images and how this factor can affect the performance. There was also no discussion on the other factors that affect the performance.
- There was a clear research gap to analysis the sufficient number of used LR images. No discussion on the adequate number of LR images and how this parameter can influence on the SR performance. Again, no discussion on the other parameters that influence on the SR performance.
- There was a clear research gap in application the wavelet-based approach to degraded satellite images for improving the performance by preserving the true edges while compressing the noise and errors.

- There was a research gap in application the wavelet-based SR approach on video sequences containing global motion, and are contaminated by noise for increasing the spatial resolution whilst supressing the noise and aliasing artefacts.
- There was a research gap in application the wavelet-based SR approach on practical video sequences containing local motions for increasing the spatial resolution while overtaking the local registration errors in low computational expensive.

3 Mathematical background

3.1 Image Observation Model

The image observation model describes the relationship between the original referenced HR image and the observed LR image/images. The image acquisition process in the spatial-domain involves warping, blurring and down-sampling to produce the LR image/images from the HR image. Figure 3-1 shows a block diagram of the observation model. Assuming that the HR image can be represented in the vector form as $x = \left[x_1, x_2, \dots, x_{L_1N_1 \times L_2N_2}\right]^T$, where $L_1N_1 \times L_2N_2$ is the size of the HR image. Assuming that L_1 and L_2 represent the down-sampling factors in the horizontal and vertical directions, respectively, and each observed LR image has the size of $N_1 \times N_2$.; the LR image can thus be denoted in the vector form by $y_k = \left[y_{k1}, y_{k2}, \dots, y_{k(N_1 \times N_2)}\right]^T$, $k = 1,2, \dots, p$, where p is the number of LR images. Assuming that each observed LR image is corrupted by additive noise, the observation model can be represented as

$$y_k = DB_kM_kx + n_k \tag{3-1}$$

where M_k is the warp matrix of size $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, B_k is the camera blur matrix with the same size, D is the down-sampling matrix of size $N_1N_2 \times L_1N_1L_2N_2$, and n_k represents the $N_1N_2 \times 1$ noise vector. It is assumed that all LR images have the same blurring model, and thus the matrix B_k can be substituted by B. These operations can be incorporated into one matrix [1], [7] and be expressed as

$$y_k = DBM_kx + n_k = H_kx + n_k \tag{3-2}$$

The matrix B generates blurred images from the warped HR image. The matrix D produces down-sampled (aliased) LR images from the warped and blurred HR image. There are various sources of blur, which are dependent on the characteristics of the main components of the camera used in the image acquisition. The first main component of the camera to add blur is the lens; because of the imperfection of the lens. Another main component of the camera that contributes in blurring the image is the image sensor; due to the insufficient sampling rate of the LR sensor and the finiteness of a physical dimension in LR sensors. The overall blur introduced by the camera is characterised via the point-spread function (PSF). Other sources of blur are the relative motion between the camera and the observed scene, which is called motion blur; and atmospheric conditions can also participate to blurring of an image [1], [15].

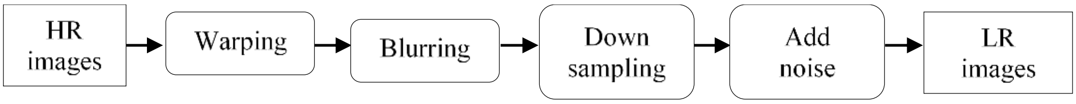
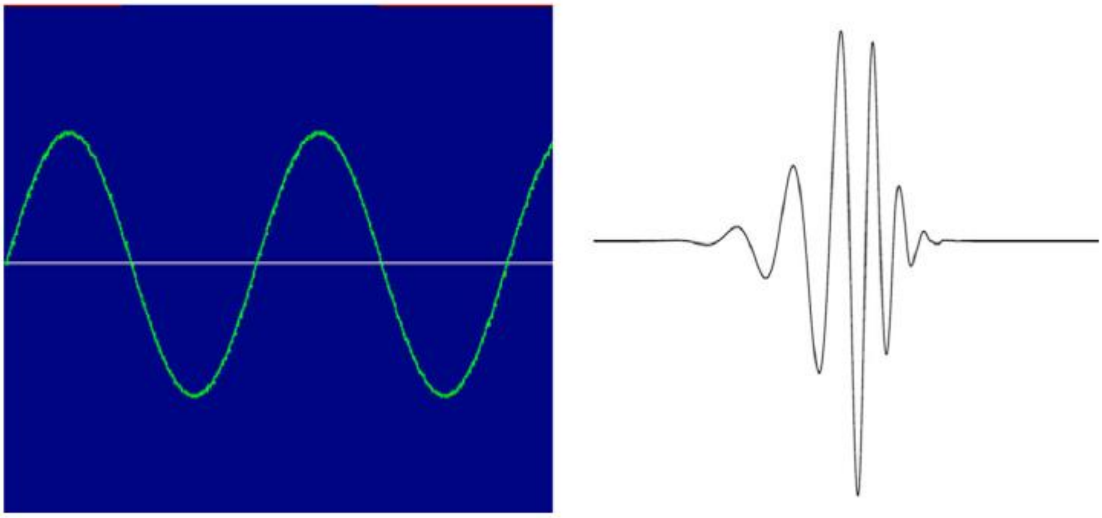


Figure 3-1 The image observation model

3.2 Wavelet Transform

Since the late 1950s, the Fourier transform (FT) has been the foundation of transform-based image processing. However, a powerful new transform, called the wavelet transform (WT), has advantages that make it more effective than the Fourier transform for compression, transmitting, and analysing many images [19]. The wavelet transform has also advantages over the Fourier transform in analysing real-wold signals where the signal contains discontinuities because the WT has attractive primary properties, such as locality, and multi-resolution [12], [16]. Different from the Fourier transform, which has bases functions called sinusoids, the wavelet transform depends upon bases functions called wavelets, of varying frequency and finite duration. This permits it to provide frequency information as well as temporal information; whereas, the Fourier transform provides only frequency information in the transformation process [19]. Wavelets are defined as functions generated from a unique mother wavelet function ∅ by dilations and translations. A wavelet is a small wave with an oscillating waveform of limited duration. The main advantage of a wavelet transform is the capability to perform local analysis. Figure 3-2 illustrates (a) sine wave and (b) wavelet of db.10. It makes sense that sharp changes can be better analysed with an irregular wavelet than a smooth sinusoid. It is also better to describe local features with wavelets which own local extent [127].



(a) (b)

Figure 3-2 (a) Sine wave and (b) wavelet of db.10 [127]

A multi-resolution analysis (MRA) approach is a new mathematical tool revealed by Mallat [128], which is beneficial in the use of wavelet bases in image analysis and leads to a rapid computation. Multi-resolution representation is very effective when analysing the information content of an image at more than one resolution [129]. The advantage of this approach is that objects of small size or low contrast are analysed at fine resolutions, while objects of large size or high contrast are examined at coarse resolution. There are two functions in multi-resolution theory: a scaling function φ and a mother wavelet function $\emptyset$. The scaling function is employed to provide a series of approximations of a signal or image. The wavelet functions are employed to add the difference in information between the neighbouring approximations [19], [129].

The wavelet transform is an effective tool that decomposes an image into low and high-frequency sub-bands locally, and then examines each sub-band with a resolution corresponded to its scale [17]. The wavelet transform can be categorised as a discrete wavelet transform (DWT), a dual-tree complex wavelet transform (DT-CWT), and a stationary wavelet transform (SWT). A one-level DWT and SWT of an input image result in one low-frequency sub-band and three high-frequency sub-bands. The high-frequency sub-bands are formed by direction filters with high magnitude responses matching to image features oriented at 0°, +45°, and 90°. The DWT is shift-variant because of the down-sampling operation exploited in the transform and, thus, each sub-band has half the size of the input image; the SWT does not cause down-sampling and, as a result, each sub-band has the same size as the input image. On the other hand, DT-CWT of the input image produces two complex-valued low-frequency sub-bands and six complex-valued high-frequency sub-bands at each level of decomposition. These high-frequency sub-bands correspond to the directions oriented at +75°, +45°, +15°, -15°, -45°, and -75° [77]. DT-CWT is approximately shift-invariant and, each sub-band has the same size as the input image. All these wavelet transforms are affected by the selection of wavelet basis. An improper wavelet basis applied in the wavelet-based methods will directly affect the accuracy of the performance of these methods.

3.3 Wavelet Bases

There are two types of wavelet bases: *orthogonal* and *biorthogonal*. Biorthogonal wavelets have advantages over orthogonal wavelets. The main advantage of biorthogonal wavelets is that they allow the use of a much wider class of filters called symmetric filters, which eliminate the problem of border effects; however they do result in coefficient expansion problem. Nevertheless, biorthogonal wavelets also have the advantage that they can use linear wavelet phase filters, which solve the problem of coefficient expansion. However, the limitation of biorthogonal wavelets is that they do not have the property of energy conservation. On the other hand, orthogonal wavelets have the advantage that they are energy preserving, but they do not allow the use of linear phase filters, except for one set of linear phase filters: Haar filters. The fact that biorthogonal wavelets are not energy conserving does not cause a significant problem; because there are linear phase biorthogonal wavelets which are very close to being orthogonal wavelets. The *Dubechies db.9/7* wavelet filter has this beneficial property [130], [131]. The *biorthogonal Daubechies 9/7* filter (also called Chohen- Daubechies-Feauveau CDF filter) is the most commonly used wavelet function for the DWT process, especially in image resolution enhancement and compression. However, it has irrational coefficients, and thus its hardware implementation requires large resources [132].

Daubechies is a well-known family of wavelets referred to as Chohen-Daubechies-Feauveau biorthogonal wavelets. As the scaling and wavelet functions of the family are symmetrical and have similar length, Daubechies wavelets are among the most widely used biorthogonal wavelets. *Symlets* (the shortened name for symmetrical wavelets) is another well-known family of wavelets. While symlets wavelets are not exactly symmetrical, they are made to have the least asymmetry [19]. Table 3-1 lists the wavelet families and their names as well as number of included filters [133].

Table 3-1 Summary of different wavelet families, their shortened names, and number of included wavelets

Wavelet Families	Shortened Family Name	Number of Wavelets	Type
Haar	'haar'	'haar' or 'db1'	Orthogonal
Daubechies	'db'	'db2', 'db3',…,'db45'	Biorthogonal
Symlets	'sym'	'sym2', 'sym3',…,'sym45'	Biorthogonal
Coiflets	'coif'	'coif1', 'coif2',…,' coif5'	Biorthogonal
Biorthogonal	'bior'	'bior1.1', 'bior1.3', 'bior1.5', 'bior2.2','bior2.4',…,'bior2.8', 'bior3.1', 'bior3.3',…, 'bior3.9', 'bior4.4', 'bior5.5', 'bior6.8'	Biorthogonal

3.4 Discrete Wavelet Transform (DWT)

The discrete wavelet transform DWT is an efficient mathematical tool for representing and analysing digital images at multiple resolutions. The DWT provides a powerful insight into an image space and frequency attributes [133]. The DWT provides suitable information for local analysis and synthesis of a time series data, or an image [12]. Firstly, it will be expressed wavelet series expansion. The generalised wavelet series expansion of function $f(x) \in L^2(R)$

can be represented by a scaling function expansion $\varphi(x)$ and wavelet function expansions $\emptyset(x)$, which can be written as

$$f(x) = \sum_{k} c_{j0}(k)\, \varphi_{j0,k}(x) + \sum_{j=j0}^{\infty} \sum_{k} d_j(k) \emptyset_{j,k}(x) \tag{3-3}$$

where j_0 is a starting scale, j is a higher scale and k is an index of a finite or infinite summation; $c_{j0}(k)$ are referred to as approximation and/or scaling coefficients and $d_j(k)$ are called detail and/or wavelet coefficients. The scaling function is employed to provide a series of approximations of $f(x)$ at scale j_0 and wavelet functions are then employed to add the difference in information between the neighbouring approximations for scale $j \geq j_0$. If the expansion scaling and wavelet functions $\varphi_{j0,k}$ and $\emptyset_{j,k}$ are an orthonormal wavelet basis, the expansion coefficients are computed by inner products of the expanded function $f(x)$ and these functions, i.e.

$$c_{j_0}(k) = \langle f(x) \varphi_{j_{0,k}}(x) \rangle = \int f(x)\, \varphi_{j_{0,k}}(x) dx \tag{3-4}$$

$$d_j(k) = \langle f(x) \emptyset_{j,k}(x) \rangle = \int f(x) \emptyset_{j,k}(x) dx \tag{3-5}$$

If the expanded signal is discrete, the coefficients are referred to as the discrete wavelet transform DWT. The one dimension (1-D) DWT decomposes an input signal $f(x)$ into a translated and dilated mother wavelet function $\emptyset(x)$ and scaling function $\varphi(x)$. For example, considering a sequence $f(n) = f(x_0 + n\Delta x)$ for $x_0, \Delta x, and\; n = 0,1,2, \ldots, M - 1$, the expansion coefficients of $f(x)$ for Eqs. (3-4) and (3-5) become the forward DWT coefficients of $f(n)$:

$$w_{\varphi}(j_0, k) = \frac{1}{\sqrt{M}} \sum_{n} f(n) \varphi_{j0.k}(n) \tag{3-6}$$

$$w_{\emptyset}(j,k) = \frac{1}{\sqrt{M}}\sum_{n} f(n)\emptyset_{j.k}(n) \dots for\ j \geq j_0 \tag{3-7}$$

where $\varphi_{j0,k}(n)$ and $\emptyset_{j,k}(n)$ are discretised forms of the bases functions $\varphi_{j0,k}(x)$ and $\emptyset_{j,k}(x)$. According to Eq. (3-3), the inverse DWT is

$$f(n) = \frac{1}{\sqrt{M}}\sum_{k} w_{\varphi}\,(j_0,k)\varphi_{j0,k}(n) + \frac{1}{\sqrt{M}}\sum_{j=j_0}^{\infty}\sum_{k} w_{\emptyset}\,(j,k)\emptyset_{j,k}(n) \tag{3-8}$$

It is assumed $j_0 = 0$ and M is chosen to be $M = 2^J$, and the summations in Eqs. (3-6) to (3-8) are applied over $n = 0,1,2,\dots,M-1, j = 0,1,2,\dots,J-1, and\ k = 0,1,2,\dots 2^j - 1$. $\frac{1}{\sqrt{M}}$ is a normalising factor added to both the forward and inverse DWT [19].

The 1-D DWT can be extended to the 2-D DWT using separable 2-D scaling and wavelet functions. Similar to the 1-D DWT, the 2-D DWT decomposes an input image $f(x,y)$ through a series of translations and dilations of a scaling function $\varphi(x,y)$ and three directional wavelet functions $\emptyset^H(x,y)$, $\emptyset^V(x,y)$, and $\emptyset^D(x,y)$. The wavelet functions provide the directional selectivity along horizontal, vertical, and diagonal directions. The DWT of $f(x,y)$ with the size of $M \times N$ is

$$W_{\varphi}(j0,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\,\varphi_{j0,m,n}(x,y) \tag{3-9}$$

$$W_{\emptyset}^{i}(j,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\,\emptyset_{j,m,n}^{i}(x,y) \tag{3-10}$$

where i refers to the index of directional wavelets. The coefficients $W_{\varphi}(j0,m,n)$ provide an approximation of $f(x,y)$ at scale $j0$ and the coefficients $W_{\emptyset}(j,m,n)$ provide the directional details for scale $\geq j_0$.

$$f(x,y) = \frac{1}{\sqrt{MN}} \sum_{m} \sum_{n} W_{\varphi}(j0,m,n)\varphi_{j0,m,n}(x,y) + \frac{1}{\sqrt{MN}} \sum_{i=H,V,D} \sum_{j=j_0} \sum_{m} \sum_{n} W_{\emptyset}(j,m,n)\emptyset_{j,m,n}(x,y) \tag{3-11}$$

A one-level DWT can be represented by filtering an input signal through low-pass and high-pass filter banks, and then down-sampling by a factor of 2 to eliminate the redundant information. Multiples scale levels are constituted by repeating the filtering and down-sampling operations on the low-pass outputs [131], [132]. A block diagram of DWT filter banks and down-samplers of level 1 are illustrated in Figure 3-3.

With separable filter banks, the 2-D DWT of an input image can be implemented by first applying the 1-D wavelet transform to all the rows of the image, and then repeating this step in all the columns. This process decomposes the input image into four filtered sub-bands: known as the approximation or low-low (LL); vertical detail or low-high (LH); horizontal detail or high-low (HL); and diagonal detail or high-high (HH) [19].

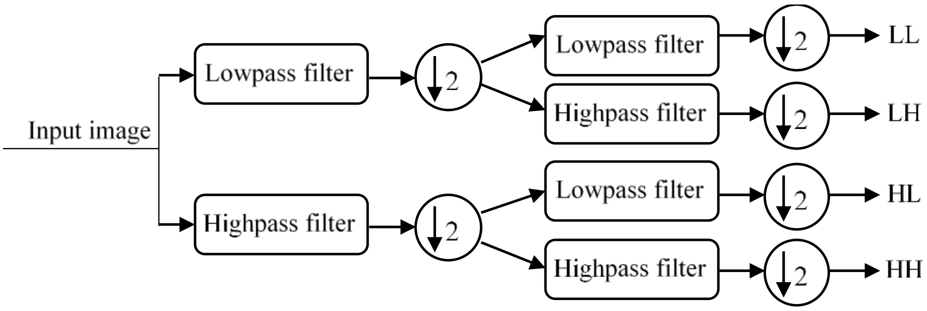


Figure 3-3 Block diagram of DWT filter banks of level 1

3.5 Dual-Tree Complex Wavelet Transform (DT-CWT)

The dual-tree complex wavelet transform DT-CWT is constructed by using two real-valued decimated DWTs. The 1-D DT-CWT separates an input signal $f(x)$ into a complex translated and dilated mother wavelet function $\emptyset(x)$ and scaling function $\varphi(x)$, i.e.

$$f(x) = \sum_{l\in z} s_{j0,l}\varphi_{j0,l}(x) + \sum_{j\geq j0} \sum_{l\in z} c_{j,l}\emptyset_{j,l}(x) \tag{3-12}$$

where z is the set of natural numbers, j and l denote the integer index of translations and dilations, respectively, $s_{j0,l}$ are the complex scaling coefficients, $c_{j,l}$ are the complex wavelet coefficients, $\varphi_{j0,l}(x) = \varphi_{j0,l}^{r}(x) + \sqrt{-1}\varphi_{j0,l}^{i}(x)$ and $\emptyset_{j,l}(x) = \emptyset_{j,l}^{r}(x) + \sqrt{-1}\emptyset_{j,l}^{i}(x)$, r and i refer to the real and imaginary parts, respectively, where the real and imaginary parts are calculated using separable filter banks.

The 2-D DT-CWT separates an input image $f(x, y)$ using a series of translations and dilations of a one complex scaling function and six complex wavelet functions, i.e.

$$f(x,y) = \sum_{l \in z^2} s_{j0,l}\, \varphi_{j0,l}(x,y) + \sum_{\theta \in \Theta} \sum_{j \geq j0} \sum_{l \in z^2} c_{j,l}^{\theta}\, \emptyset_{j,l}^{\theta}(x,y) \tag{3-13}$$

Here $\theta \in \Theta = \{\pm 15^{\circ}, \pm 45^{\circ}, \pm 75^{\circ}\}$ creates the direction of the complex wavelet function; in other words, a one-level DT-CWT of the input image results in one complex-valued low-frequency sub-band and six complex-valued high-frequency sub-bands [65], [134]. The advantages of DT-CWT is that it has a shift-invariant property and good directional selectivity, and benefits from using direction-selective filters with high magnitude responses in the presence of image features.

3.6 Image Interpolation

Interpolation is the widely used technique for increasing the resolution of a digital image. Fundamentally, *image interpolation* is the process of using known data to estimate values at unknown locations [19]. Interpolation is a basic tool used extensively for several different applications in image processing, including zooming, shrinking, rotating, and geometric corrections. The sinc interpolation function produces an exact reconstruction of a continuous function, if the data was sampled at or above the Nyquist rate [79]. However, the sinc basis function does not yield satisfactory results for image data as image data is acquired at a much lower sampling rate [44]. The classical polynomial interpolation methods, such as Lagrange interpolation, also do not produce acceptable results as the global polynomial model is not able to precisely model local image features. The piecewise polynomial approaches, which are also called splines, were investigated to minimise the limitations suffered by the global polynomial methods [79]. There are four well-known interpolation methods: nearest neighbour, bilinear, bicubic, and Lanczos.

Nearest neighbour interpolation is the easiest method, where the intensity value at the new location is assigned as the value at the old location, which is the nearest neighbour to the new location. The main drawback of this method is that it results in staircase artefacts. An improved method is bilinear interpolation, in which the value of the new location is estimated using the four nearest neighbour pixels by taking a weighted average of these pixels. This method produces much better results than the nearest neighbour interpolation, with a moderate increase in computational cost [19], [81], [82]. Bicubic interpolation is a more complex method, which is also referred to as cubic convolution interpolation in [80] and HR cubic spline interpolation in [81], [82], and [83]. In this method, the intensity value of the new location is assigned using the sixteen nearest neighbour points by taking a weighted average of these points. This method is more efficient and accurate than bilinear interpolation and nearest neighbour, by producing sharper edges in terms of preserving more of the details in the original image [80]: it is the most preferred interpolation method in commercial image editing programs [19]. Lanczos interpolation is a windowed form of the sinc function. There are two versions of such a windowed sinc function: the Lanczos windowed sinc function of degree 2 and 3. This method also produces good results in terms of its capability to detect edges and linear features. It gives a good compromise in terms of reduction of aliasing and ringing artefacts, and sharpness of edges [67], [84]. The main drawback of most interpolation methods is that the images produced suffer from blurring and staircase artefacts.

3.7 Summary

This chapter presents the basic concepts of wavelet-based image resolution enhancement techniques, and SR techniques. Although the goal of SR reconstruction methods is to produce one, or a set of, HR images from multiple

LR images, it also covers resolution enhancement methods that produce a single HR image from a single LR image. The image observation model is the first step to analyse resolution enhancement and SR reconstruction processes, which is applicable for degradation of both still images and video sequences. The second step is to describe the wavelet-domain, which is the employed domain. This chapter contains the key elements of the wavelet transform, which include the wavelet bases and types of wavelet transforms.

4 An Optimal Factor Analysis Approach to Improve the Wavelet-based Image Resolution Enhancement Techniques

4.1 Introduction

This chapter studies the existing wavelet-based image resolution enhancement techniques in terms of the way to assess their performance. It has been observed that the existing wavelet-based methods have many assumptions of the factors that affect the performance. A major limitation of most these methods is that the assumptions they make are not always satisfied for real applications, which limit their performance in practice. For example, the detail of a physical object that an optical instrument can reproduce in an image has limits which are mandated by the laws of physics, whether formulated by the diffraction equations in the wave theory of light or the Uncertainty Principle for photons in quantum mechanics. There is no such a well-accepted model that can fully describe the underlying mechanism to produce the observed LR image. This mechanism can also be varied case by case. Another example is that the selection of different wavelet functions is either very limited or never considered. Most of the existing wavelet-based image resolution enhancement methods evaluate their performance using the wavelet function db.9/7. Although the db.9/7 wavelet function is the most widely used one in the wavelet-based resolution enhancement methods, there is a wide range of wavelet families and their different wavelet functions for chosen. It has also been observed the inconsistency of assumptions of the factors for each method. For example, the studied methods assume that the observed LR image is obtained either from down-sampling the original HR image by applying discrete wavelet transform (DWT) and considering a low-frequency (LL) sub-band of DWT, or low-pass filtering and then down-sampling. In other words, the superior of one method than other methods claimed in the literatures is conditional. Although it has been reported in [135] that the performance of resolution enhancement techniques can be affected by the methods to produce the LR images, and other factors, there is very limited literatures investigating how to utilise these factors to assess and improve the performance of these techniques.

Resolution enhancement techniques in the wavelet-domain have attracted more and more investigations to address the problems associated with conventional interpolation methods. Wavelet-based resolution enhancement methods aim to reconstruct the HR image from the given LR image by estimating the unknown wavelet coefficients in sub-bands containing high-pass frequency information. A simple approach, called wavelet-zero padding (WZP), is to recover an approximation of the HR image by filling the unknown wavelet coefficients with zeros and applying the inverse wavelet transform [62]. Although, this method is able to surpass the conventional interpolation methods, it commonly introduces artefacts, such as smoothing and ringing, into the reconstructed HR image. Many advanced methods have been introduced to estimate the wavelet coefficients of high-frequency sub-bands. Hidden Markov tree (HMT) approach in the wavelet- domain is able to exploit the statistical relationships between wavelet coefficients at different scales. A statistical framework for signal processing was introduced by Crouse et al. [16] using wavelet-domain hidden Markov models (HMMs). A hidden Markov tree -based method using HMM approach [16] was proposed by Kinebuchi et al. in [58] and an extended version of HMM approach [16] was presented by Zhao et al. in [59], to estimate the unknown detail coefficients by the mixed Gaussian distributions, which are symmetrical around zero. However, the performance of these methods is affected by the sign changes between the scales not being taken into account. To reduce this drawback, a refined HMT-based method was proposed by Temizel in [60], where the coefficient magnitude estimation is separated from the sign estimation. A non-directional cycle-spinning (CS) method [62], called WZP-CS, was developed by Temizel and Vlachos as an effective method for reducing ringing artefacts by averaging out the translated zero-padded reconstructed images. However, ringing artefacts not only occur at the vicinity of edges; in particular, they are predominantly correlated with the orientation of edges. A directional CS method, called WZP-DCS, was introduced in [63] by Vlachos, which can refine better edge orientation and prevent ringing artefacts.

A further improvement of this method could be obtained by applying a CS and edge rectification technique [64]. A dual-tree complex wavelet transform (DT-CWT) approach was introduced by Reeves and Kingsbury [65]. Recently, DT-CWT technique [66] was proposed by Demirel and Anbarjafari for resolution enhancement of satellite images. One-level DT-CWT decomposes an input LR image into different frequency sub-bands, and then the high-frequency sub-bands produced by DT-CWT and the input image are interpolated using bicubic interpolation. Finally, a super-resolved image is generated by combining all these interpolated sub-bands through the inverse DT-CWT. More recently, a DT-CWT technique based on non-local means filter and Lanczos interpolation was proposed by Iqbal et al. [67] for resolution enhancement of satellite images. In recent years, discrete wavelet transform DWT-based image resolution enhancement techniques have attracted increasing investigations. A Demirel-Anbarjafari Super Resolution (DASR) method [70] was proposed, in which the decomposed three high-frequency sub-bands and the input image are interpolated using bicubic interpolation. Upgrading from the DASR, a DWT-Diff technique [71] was introduced by Demirel and Anbarjafari with its application in satellite images, in which the high-frequency sub-bands are further enhanced by considering the difference between the input LR image and the interpolated low-frequency sub-band. The same authors proposed a method based on DWT- and stationary wavelet transform (SWT) [72], which introduced an intermediate process by adding the high-frequency sub-bands obtained through SWT of the input image with the high-frequency sub-bands obtained through DWT.

4.2 State-of-The-Art Image Resolution Enhancement Methods

In this section, the wavelet zero padding (WZP) and cycle-spinning (CS) methods will be described below as an example of the state-of-the-art wavelet-based image resolution enhancement methods, aiming to use for performance assessment because they are relatively simple.

4.2.1 Wavelet Zero Padding (WZP)

Wavelet zero padding WZP is the simplest method for image resolution enhancement in the wavelet-domain. An initial approximation to the unknown HR image is obtained by using an input LR image as low-frequency sub-band and zero-padding of high-frequency sub-bands (i.e. filling all the elements of these sub-bands with zeros), and applying the inverse wavelet transform (IWT). A simplified block diagram of the method is shown in Figure 4-1.

$$\hat{x}_0 = w^{-1}\begin{bmatrix} y & 0_{m,n} \\ 0_{m,n} & 0_{m,n} \end{bmatrix} \tag{4-1}$$

where $\hat{x}_0$ is an initial approximation to the unknown HR image, y is an input LR image of size $m \times n$, $0_{m,n}$ represents an all-zero sub-matrix of dimensions $m \times n$, and w^{-1} is the inverse discrete wavelet transform. This method is relatively simple to implement and is capable of outperforming the conventional interpolation methods [62].

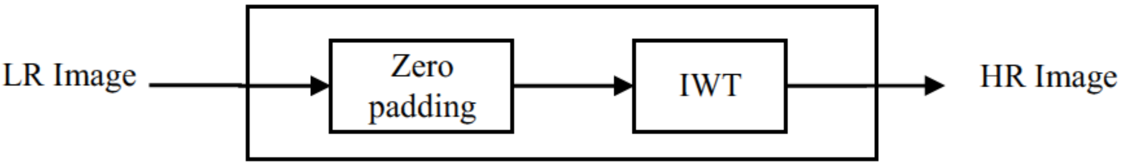


Figure 4-1 A simplified block diagram of the WZP method [62]

4.2.2 Cycle-Spinning (CS)

A non-directional cycle-spinning CS method, called WZP-CS, was used as an effective tool toward reducing ringing artefacts. A non-directional CS technique in the wavelet-domain can be summarized by the following steps:

1) An initial approximation to the unknown HR image $\hat{x}_0$ is obtained using WZP.
2) A number of LR images $\hat{y}_{i,j}$ are generated from the initial HR approximation image in step (1) by spatial horizontal and vertical shifts, wavelet transform, and discarding the high-frequency sub-bands.

$$\hat{y}_{i,j} = DWS_{i,j}\hat{x}_0 \qquad \textbf{(4-2)}$$

where D denotes discarding of high-frequency sub-bands, W is wavelet transform, $S_{i,j}$ represents a shift operator by applying horizontal and vertical shifts of $i, j \in \{-k, k+1, \dots, k-1, k\}$., where $k \in \{1,2,3,\dots,N\}$.

3) WZP processing is applied to all those LR images $\hat{y}_{i,j}$ to give N HR images $\hat{x}_{i,j}$, where $N = (2k+1)(2k+1)$. These intermediate HR images are realigned and averaged to give the final reconstructed HR image $\hat{x}$, as written by Eq. (4-3), where $S_{i,j}^{-1}$ denotes the inverse of the shifting operator [62].

$$\hat{x} = \frac{1}{N}\sum_{i=-k}^{k}\sum_{j=-k}^{k} S_{i,j}^{-1}\hat{x}_{i,j} \qquad \textbf{(4-3)}$$

A block diagram of the algorithm is illustrated in Figure 4-2.

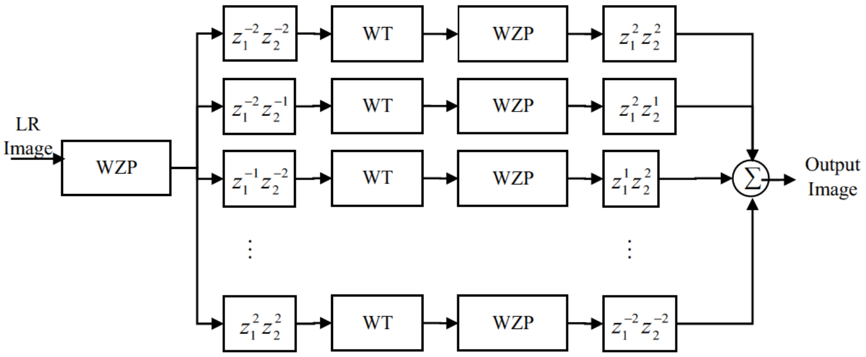


Figure 4-2 Block diagram of the WZP-CS method [62]

4.3 Proposed Methods

4.3.1 Important Factors

Table 4-1 summaries the reviewed wavelet-based image resolution enhancement techniques in terms of the way to evaluate their performance. The inconsistencies in the assumptions of the considered factors for each individual technique has been observed.

Table 4-1 Summary of different wavelet-based image resolution enhancement techniques in terms of performance assessment.

Techniques	Input LR Image	Scale Factor	Interpolation Method	Wavelet Function	Test Image
WZP-CS [62]	LL sub-band of DWT	2 & 4	N/A	Db.9/7	Lena, Elaine, Baboon, and Peppers
WZP-DCS [63]	Low-pass filtering and down-sampling	2 & 4	N/A	Db.9/7	Lena, Elaine, Baboon, and Peppers
HMT [58]	Down-sampling of HR image	N/A	N/A	N/A	Lena
HMT [59]	Down-sampling of HR image	2	N/A	N/A	Lena
HMT [60]	Low-pass filtering and down-sampling	2 & 4	N/A	Db.9/7	Lena, Elaine, Baboon, and Peppers
DT-CWT [66]	LL sub-band of DWT	2 & 4	Bicubic	N/A	5 Satellite Images
DT-CWT [67]	Down-sampling of HR image	4	Lanczoc	N/A	1 Satellite Image (Washington DC)
DASR [70]	LL sub-band of DWT	4	Bicubic	Db.9/7	Lena, Elaine, Baboon, and Peppers
DWT-Diff [71]	LL sub-band of DWT	4	Bicubic	Db.9/7	5 Satellite Images
DWT-SWT [72]	Down-sampling of HR image	4	Bicubic	Db.9/7	Lena, Elaine, Baboon, and Peppers
DWT-SWT [73]	Down-sampling of HR image	4	Bicubic	N/A	Lena, Elaine, Head, and Brain

For example, the considered methods make the assumption that the observed LR image is produced by either applying a low-pass filtering and then downsampling, or achieving the low-frequency LL sub-band of DWT. For some methods, the description of these factors is either neglected or unclear. The performance of these methods is unknown when such an assumption is not satisfied. A method to compare the resolution enhancement methods in a more comprehensive and equitable way is therefore required. Such a method can also be used to further improve the overall performance of the existing methods. Each potential factor that affects the performance has been studied one by one.

4.3.1.1 The Mechanism to Produce Low-Resolution Images

It has been identified from the literature review that there are various ways to generate LR images including down-sampling of the original HR image through (a) DWT by the db.9/7 wavelet function or by the Haar wavelet function, (b) bicubic interpolation, (c) bilinear interpolation, (d) nearest neighbour, and (e) low-pass filtering and then down-sampling. Table 4-2 shows the resulting PSNR values for Lena image using different resolution enhancement methods by considering different LR image generation methods.

Table 4-2 PSNR results for Lena image using different techniques for resolution enhancement from 128×128 to 512×512 for several LR image generation methods.

Techniques	PSNR (dB)					
	DWT by DB.9/7	DWT by Haar	Bicubic	Bilinear	Nearest	Low-pass
WZP(haar)	22.36	25.77	25.75	25.19	24.35	25.18
WZP(db.9/7)	**24.22**	25.75	25.73	25.23	23.21	24.04
Bicubic	22.51	**26.31**	**26.28**	**25.75**	24.80	**25.67**
Bilinear	22.63	25.53	25.54	24.85	**24.87**	25.21
Nearest	21.53	24.71	24.61	24.44	22.79	23.97

Inspection of Table 4-2 shows that WZP method with the wavelet function db.9/7 has the best performance among the considered methods for the input LR image produced by DWT with db.9/7. For the LR images obtained by DWT with Haar, bicubic interpolation, bilinear interpolation and low-pass filtering methods, the bicubic interpolation method has the highest PSNR values, but for the LR image produced by the nearest neighbor, the bilinear interpolation method has the best performance. These observations clearly indicate that the method to produce LR image has significant effect on the performance of different techniques.

4.3.1.2 Wavelet Function

There are several well-known wavelet families such as Daubechies (db), Symlets (sym), Biorthogonal (bior), Coiflets (coif), etc. [133]. In this chapter, the behaviour of the considered resolution enhancement techniques has been studied for a wide range of wavelet families as well as their various wavelet functions, including (db1-20), (sym2-20), (bior1.1-6.8) and (coif1-5). Note that db.1 is also referred to as Haar, and db.9/7 is equivalent to bior4.4 [129]. Table 4-3 illustrates the PSNR values for three well-known test images (Lena, Baboon, and Elaine) using the WZP method with various wavelet functions, where only the wavelet functions producing high PSNR values are shown to save space. The input LR image has been produced by down-sampling the original HR image using DWT with db.9/7 wavelet function. The quantitative results show that coif2, sym3, and db3 are top three wavelet functions in terms of PSNR values for all three test images, not the well investigated Haar or bior4.4. This observation indicates that the selection of wavelet function can play a key role in improving the performance. However, in most of the existing wavelet-based resolution enhancement methods, the discussion of selection different wavelet functions is very limited.

Table 4-3 PSNR results for three well-known test images (Lena, Baboon, and Elaine) generated by DWT with db.9/7 using different techniques for resolution enhancement from 128×128 to 512×512 of various wavelet families and functions.

Techniques	PSNR (dB)		
	Lena	Baboon	Elaine
Bicubic	22.51	24.21	25.49
Bilinear	22.63	24.23	25.52
Nerest	21.53	23.49	24.40
WZP (haar)	22.36	24.09	25.31
WZP (bior 1.1)	22.36	24.09	25.31
WZP (bior 2.2)	24.19	25.19	27.40
WZP (bior 3.1)	22.64	24.27	25.57
WZP (bior 4.4)	24.22	25.23	27.46
WZP (bior 5.5)	24.13	25.19	27.41
WZP (bior 6.8)	24.22	25.22	27.44
WZP (sym2)	25.32	25.67	28.77
WZP (sym3)	**26.45**	**26.26**	**30.15**
WZP(sym7)	25.59	25.91	28.90

WZP (sym15)	25.56	25.88	29.08
WZP (sym19)	26.15	26.13	29.86
WZP (coif1)	24.16	25.18	27.39
WZP (coif2)	**26.56**	**26.30**	**30.28**
WZP (coif3)	24.08	25.16	27.41
WZP (db2)	25.32	25.76	28.77
WZP (db3)	**26.45**	**26.26**	**30.15**
WZP (db4)	24.21	25.23	27.51

4.3.1.3 Enlargement Factor

As shown in Table 4-1, the performance of most methods has been evaluated by an enlargement factor of 2 or 4. To better evaluate the effectiveness of this factor on performance, this study has considered different scale factors of 2, 4, 8 and 16. The input LR image has been obtained by down-sampling using DWT with db.9/7 wavelet function. The produced PSNR values for Lena image are shown in Table 4-4, inspection of which shows that WZP method with db.9/7 produces the highest PSNR values for all enlargement factors. The Lanczos and bicubic techniques provide higher PSNR values than bilinear technique for the scale factor of 2 but for the scale factors of 4, 8 and 16, the bilinear technique produces higher PSNR values than Lanczos and bicubic techniques. The performance variation of the considered methods decreases following the increase of scale factor, which indicates that the scale factor is an important parameter to be considered for performance assessment.

Table 4-4 PSNR results for Lena image generated using DWT with db.9/7 for enlargement factors of 2, 4, 8 and 16 using different techniques.

Techniques	PSNR (dB)			
	Factor 2	Factor 4	Factor 8	Factor 16
WZP(db.9/7)	**32.93**	**24.22**	**19.89**	**17.22**
WZP(haar)	26.44	22.36	19.26	16.97
Bicubic	28.05	22.51	19.28	16.98
Lanczos	28.06	22.39	19.16	16.83
Bilinear	27.77	22.63	19.46	17.22
Nearest	26.44	21.53	18.60	16.32

4.3.1.4 Interpolation Function

Because of the obvious weakness of the nearest neighbor method, it has been neglected in this study, and the bilinear, bicubic and Lanczos methods have been tested. The input LR image has been produced by down-sampling using DWT with db.9/7. The PSNR results for Lena image are illustrated in Table 4-5, inspection of which indicates that there is no significant difference in performance for different interpolation methods. Moreover, the interpolation method producing the highest PSNR value is not consistent for different methods. These observations indicate that the selection of interpolation function for wavelet-based techniques can affect the performance, but not significantly.

Table 4-5 PSNR results for Lena image generated using DWT with db.9/7 for resolution enlargement factor from 128×128 to 512×512 using different resolution enhancement techniques.

Techniques	PSNR (dB)		
	Bicubic	Lanczos	Bilinear
WZP-CS(db.9/7)	**24.23**	24.18	24.05
WZP(db.9/7)	**24.22**	24.18	24.05
WZP(haar)	22.36	22.23	**22.52**
WZP(coif2)	26.56	**26.88**	25.60

4.3.1.5 Test Image

In order to show the effectiveness of the test images on the performance of the considered resolution enhancement methods, three well-known standard images (Lena, Baboon, and Elaine) with different features have been tested for comparison. Table 4-6 illustrates the PSNR values using different conventional interpolation and state-of-the-art wavelet-based resolution enhancement methods. The input image has also been generated by DWT with db.9/7 wavelet filter. Inspection of Table 4-6 shows that Elaine image with many sharp edges [72], has the highest PSNR improvement. For Lena image with many strong edges relative to the number of textures [57], the improvement is lower. However, Baboon image with many textures [72], has the lowest PSNR improvement because of the relative redundancy of textures. These observations indicate that the test image is an important factor to be considered for performance evaluation.

Table 4-6 PSNR results for three well-known test images (Lena, Baboon, and Elaine) generated by DWT with db.9/7 using different techniques for resolution enhancement from 128×128 to 512×512.

Techniques	PSNR (dB)		
	Lena	Baboon	Elaine
WZP(coif2)	26.56	26.30	**30.28**
WZP-CS(db.9/7)	24.23	25.23	**27.47**
WZP(db.9/7)	24.22	25.22	**27.46**
WZP(haar)	22.36	24.09	**25.31**
Bicubic	22.51	24.21	**25.49**
Lancozos	22.39	24.16	**25.43**
Bilinear	22.63	24.23	**25.52**

4.3.2 Image Fidelity Criteria

There are two types of criteria for assessing the performance of resolution enhancement algorithms: (1) objective fidelity criteria and (2) subjective fidelity criteria.

Objective fidelity criterion or quantitative metric evaluates the performance as a mathematical function of the input and output images. The root-mean-square error ($RMSE$) between the original HR image and the output image is one of the commonly used objective fidelity criteria. Consider $f(x,y)$ is an input image and $\hat{f}(x,y)$ is an output image. For any value of x and y, the error $e(x,y)$ between $f(x,y)$ and $\hat{f}(x,y)$ is

$$e(x,y) = \hat{f}(x,y) - f(x,y) \qquad \textbf{(4-4)}$$

The total error between the two images is

$$e_{sum} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [\hat{f}(x,y) - f(x,y)] \qquad \textbf{(4-5)}$$

where the images are with size $M \times N$. The $RMSE$ between these two images is the square root of the squared error averaged over the $M \times N$ array, and it can be expressed as

$$RMSE = \left[\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [\hat{f}(x,y) - f(x,y)]^2 \right]^{1/2} \qquad \textbf{(4-6)}$$

If the output image $\hat{f}(x,y)$ is considered to be the input image $f(x,y)$ and the error or difference is considered to be "noise" signal $e(x,y)$, the mean-square signal-to-noise ratio (SNR_{ms}) of the output image can be defined as

$$SNR_{ms} = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \hat{f}(x,y)^2}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [\hat{f}(x,y) - f(x,y)]^2} \qquad \textbf{(4-7)}$$

The ratio (SNR_{ms}) offers a measure of the level of signal power to the level of noise power. An image with low noise has a high ratio and the same image with a larger level of noise has a lower ratio.

The peak-signal-to-noise ratio $(PSNR)$ between the resolution enhanced image and the original HR image is one of the most important and commonly used quantitative measures. It can be calculated as

$$PSNR = 10 \log_{10}(\frac{L^2}{MSE}) \qquad \textbf{(4-8)}$$

where L is the maximum fluctuation in the image. If the image is represented by 8-bit grayscale, the value of L will be 255. MSE represents the mean-square-error between the resolution enhanced image and the original HR image. It can be calculated as

$$MSE = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\hat{f}(x,y) - f(x,y)\right]^2 \quad \textbf{(4-9)}$$

Although objective fidelity criteria give a simple and appropriate way to assess the performance, subjective fidelity criteria are often more convenient. It is based on evaluating image quality by the subjective assessments of humans [19].

4.3.3 Optimal Factor Analysis (OFA)

The behaviour of resolution enhancement methods has been assessed above by varying one factor and fixing other factors, which aims to identify the important factors but it cannot reveal the best technique with the optimal parameter selection. Addressing this challenge, this chapter proposes an Optimal Factors Analysis (OFA) approach in order to increase the performance of the existing wavelet-based methods, and also better assess their overall performance.

OFA algorithm considers a resolution enhancement technique, $\emptyset$, as a Multi-Input and Multi-Output (MIMO) model, which includes five inputs variables: the way to produce LR image $LR_a(a = 1,2,\dots,A)$, the scale factor $SF_b(b = 1,2,\dots,B)$, the test image $TI_c(c = 1,2,..,C)$, the wavelet function $WF_d(d = 1,2,\dots,D)$, and the interpolation method $IM_e(e = 1,2,\dots,E)$, where A, B, C, D, and E are the total number of possible states for five variables, respectively. There are three outputs including the highest $PSNR$ value $PSNR^*$, the optimal wavelet function WF^* and the optimal interpolation method IM^*. The MIMO model can therefore be written as:

$$(PSNR^*, WF^*, IM^*) = F_{\emptyset}(LR_a, SF_b, TI_c, WF_d, IM_e) \quad \textbf{(4-10)}$$

Depending on the values of A, B, C, D, and E, Eq. (4-10) can be solved by either an exhausted search or advanced optimisation techniques. In this study, the exhausted search has been employed.

The advantages of proposing the OFA approach are to find the best factors and evaluate the performance of the existing wavelet-based resolution enhancement techniques in a more comprehensive and equitable method in order to increase the applicability and fidelity of the existing methods. However, the disadvantage of OFA method is that it has to scan all possible states for input variables, e.g. it has to consider different methods to generate LR images, and different wavelet functions and different interpolation methods.

To better evaluate the overall performance, this chapter proposes a new Figure of Merit (FoM) measure, called Ratio of PSNR (RPSNR) that considers the 'bicubic' interpolation as the baseline. For a test image TI_c, a way to produce LR image LR_a, and a scale factor SF_b, RPSNR of the technique $\emptyset$ can be written as

$$RPSNR_{\emptyset}(LR_a, SF_b, TI_c) = \frac{\max\{PSNR_{\emptyset}(LR_a, SF_b, TI_c, WF_d, IM_e)\}}{PSNR_{\emptyset}(LR_a, SF_b, TI_c, 'bicubic')} \quad \textbf{(4-11)}$$

A higher $RPSNR$ indicates a better performance. To collectively assess the performance of $\emptyset$ over all considered factors, the averaged RPSNR is introduced and expressed as

$$\overline{RPSNR_{\emptyset}} = \frac{1}{A \times B \times C} \sum_{a=1}^{A} \sum_{b=1}^{B} \sum_{c=1}^{C} RPSNR_{\emptyset}(LR_a, SF_b, TI_c) \quad \textbf{(4-12)}$$

4.4 Results and Discussions

This study has considered six methods to generate input LR images ($A = 6$), including DWT with db. 9/7 wavelet function, DWT with Haar wavelet function, bicubic, bilinear, nearest, and low-pass filtering. Three scale factors 2, 4, and 8 ($B = 3$) and three tested images ($C = 3$) including Lena, Baboon, and Elaine have also been considered in this study. Considered wavelet families and their functions include Daubechies (db1 to db20), Symlets (sym2 to sym20), Coiflets (coif1 to coif5) and Biorthogonal (bior1.1 to bior6.8). Considered resolution enhancement techniques can be classified into five groups: interpolation methods and four WZP-based methods with different wavelet families (WZP+db, WZP+sym, WZP+coif, and WZP+bior). Three interpolation methods ($E = 3$) have been considered, namely bilinear, bicubic and Lanczos.

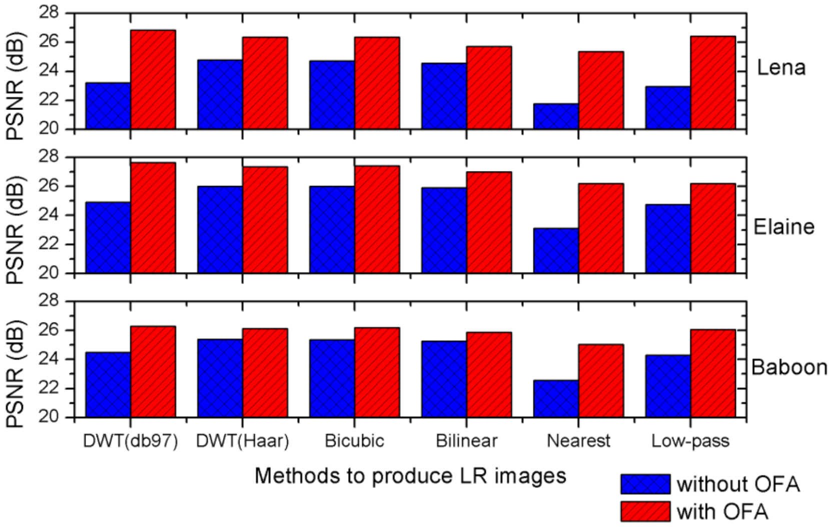


Figure 4-3 Performance improvement for the WZP techniques by applying the proposed OFA method for the scale factor of 4.

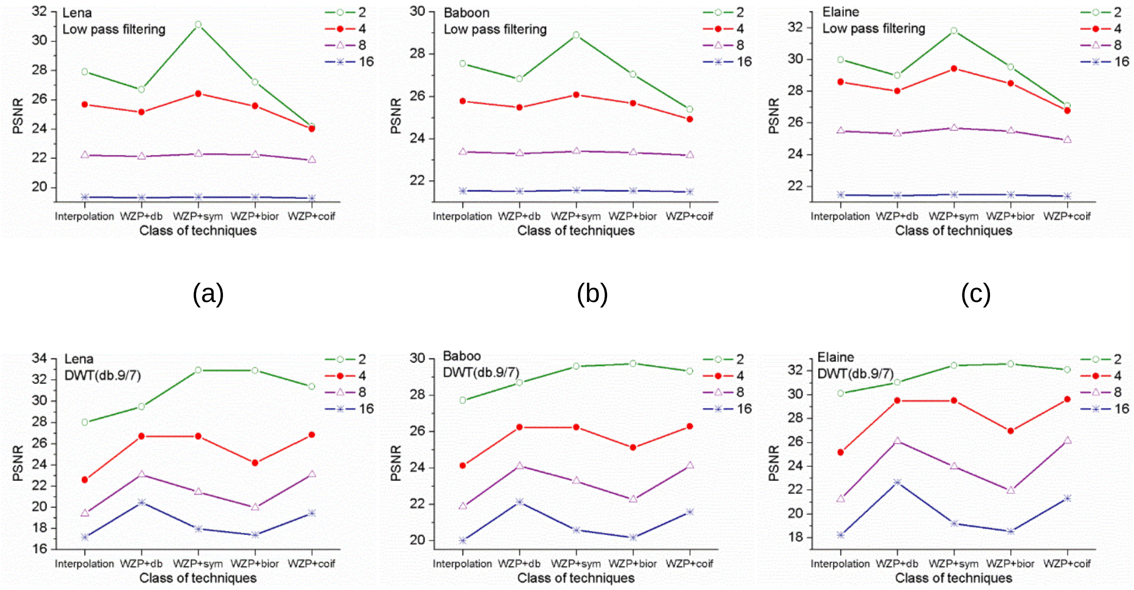
Figure 4-3 illustrates the performance of the WZP-based methods before and after applying the proposed method, where the LR images have been resolution enhanced from 128×128 to 512×512. The blue and red bars plot the PSNR values before and after applying OFA method, respectively. It is clearly shown that the proposed method significantly improves the performance of the WZP-based techniques for all six ways to produce LR images and for all three tested images.

Table 4-7 shows the results including the best performed method with its optimal parameter selection, as well as the highest PSNR and RPSNR value for different factors. For the LR image obtained from DWT with db. 9/7 wavelet function, the optimal class corresponding with the optimal interpolation method is WZP using "sym" with bilinear interpolation for the Lena image with the scale factor of 2. However, for the Baboon and the Elaine images, the best class is WZP using "bior" with bilinear interpolation for the Elaine image and with Lanczos interpolation for the Baboon image. For the scale factors of 4 and 8, the best class with the best interpolation method is WZP using "coif" with Lanczos interpolation for all three tested images.

Table 4-7 Highest PSNR and RPSNR results corresponding with optimal resolution technique and interpolation method for Lena, Baboon and Elaine images with three scale factors 2, 4, and 8.

Testing image	Scale factor	Methods to produce LR image					
		DWT by DB.9/7	DWT by Haar	Bicubic	Bilinear	Nearest	Low-pass
Lena	2	WZP(sym20) Bilinear 32.98(1.1760)	Interpolation Lanczos 32.63(1.0221)	Interpolation Lanczos 32.48(1.0237)	Interpolation Lanczos 30.85(1.0218)	WZP(sym18) Lanczos 31.22(1.1319)	WZP(sym18) Bilinear 31.18(1.1232)
	4	WZP(coif2) Lanczos 26.88(1.1942)	Interpolation Lanczos 26.56(1.0097)	Interpolation Lanczos 26.58(1.0117)	Interpolation Lanczos 25.89(1.0125)	WZP(sym18) Bicubic 25.40(1.0243)	WZP(sym18) Lanczos 26.45(1.0305)
	8	WZP(coif4) Lanczos 23.14(1.2004)	Interpolation Lanczos 23.05(1.0058)	Interpolation Lanczos 23.09(1.0078)	Interpolation Lanczos 22.64(1.0096)	WZP(sym8) Bilinear 21.76(1.0150)	WZP(sym9) Bicubic 22.35(1.0035)
Baboon	2	WZP(bior4.4) Lanczos 30.07(1.0777)	Interpolation Lanczos 29.64(1.0038)	Interpolation Lanczos 29.67(1.0094)	WZP(sym13) Bilinear 28.97(1.0106)	WZP(sym6) Bilinear 28.06(1.0419)	WZP(sym18) Bilinear 29.19(1.0582)
	4	WZP(coif2) Lanczos 26.43(1.0918)	Interpolation Lanczos 26.33(1.0029)	Interpolation Lanczos 26.39(1.0053)	Interpolation Lanczos 26.03(1.0063)	WZP(sym18) Bilinear 25.25(1.0251)	WZP(sym18) Bicubic 26.21(1.0119)
	8	WZP(coif4) Lanczos 24.22(1.1073)	Interpolation Lanczos 24.06(1.0031)	Interpolation Lanczos 24.10(1.0051)	WZP(bior5.5) Lanczos 23.86(1.0067)	WZP(bior3.1) Bilinear 22.76(1.0212)	WZP(sym6) Bilinear 23.50(1.0040)
Elaine	2	WZP(bior4.4) Bilinear 34.96(1.1177)	Interpolation Lanczos 34.54 (1.0074)	Interpolation Lanczos 34.56(1.0115)	Interpolation Lanczos 33.56(1.0122)	WZP(sym6) Bilinear 32.71(1.0700)	WZP(sym18) Lanczos 33.73(1.0901)
	4	WZP(coif2) Lanczos 30.64(1.2021)	Interpolation Lanczos 30.42 (1.0108)	Interpolation Lanczos 30.49(1.0116)	Interpolation Lanczos 29.70(1.0130)	WZP(sym18) Bicubic 29.35(1.0343)	WZP(sym18) Lanczos 30.39(1.0379)
	8	WZP(coif4) Lanczos 26.58(1.2488)	Interpolation Lanczos 26.60 (1.0130)	Interpolation Lanczos 26.63(1.0134)	Interpolation Lanczos 25.89(1.0147)	WZP(sym17) Bicubic 25.26(1.0064)	WZP(sym17) Bicubic 26.08(1.0073)

For the LR images obtained from DWT with Haar, bicubic, and bilinear, the optimal technique with the highest PSNR value is Lanczos interpolation for most of the cases. For the LR images produced by nearest and low-pass filtering, the best class is WZP using “sym” for almost all cases. These observations conclude that, for the LR images obtained from DWT with db. 9/7, nearest and low-pass filtering, the wavelet-based techniques have the biggest potential to outperform the conventional interpolation methods, due to the fact that they have relatively large RPSNR values. For the LR images produced by Haar, bicubic, and bilinear, the wavelet-based methods have no significant advantages over the interpolation methods. This conclusion justifies that for almost all papers about wavelet-based resolution enhancement techniques, the LR image has been produced by either DWT with db. 9/7 wavelet function or low-pass filtering.



(a) (b) (c)

(d) (e) (f)

Figure 4-4 Highest PSNR values for each class of resolution enhancement technique for different test images and LR image producing ways. (a) Lena with low pass filtering; (b) Baboon with low pass filtering; (c) Elaine with low pass filtering; (d) Lena by DWT with db. 9/7; (e) Baboon by DWT with db. 9/7; Elaine by DWT with db. 9/7.

In order to show the sensitivity for the selection of class of technique with different scale factors, input LR image producing methods and test images, the highest PSNR value for each class of technique has been detected and the results are shown in Figure 4-4. The standard deviation (std) for each scale factor has been calculated to describe the performance variation of each class. Table 4-8 shows the std values for the three tested images generated by low-pass filtering and DWT with db.9/7 respectively, for scale factors of 2, 4, 8 and 16. A high std value indicates that the selection of class is important because the performance for different classes of techniques is significantly varied. A low std value indicates that the performance for each class of technique is relatively similar.

Table 4-8 Standard deviation results for three test images (Lena, Baboon, and Elaine) obtained by low-pass filtering and DWT with db. 9/7 for scale factors 2, 4, 8, and 16.

Scale Factor	Lena		Baboon		Elaine	
	Low-pass	DWT by db. 9/7	Low-pass	DWT by db. 9/7	Low-pass	DWT by db. 9/7
2	2.50	2.16	1.27	0.83	1.71	1.05
4	0.88	1.92	0.43	0.99	0.97	2.01
8	0.16	1.70	0.07	1.03	0.28	2.27
16	0.03	1.40	0.03	0.92	0.04	1.91

Figure 4-4 (a), (b) and (c) illustrate the sensitivity of the class selection for Lena, Baboon, and Elaine respectively, with the LR images obtained by low-pass filtering. It is observed that if the scale factor is high, the PSNR is low as expected, and importantly the std is low. This observation means that different classes of techniques have similar performance for a larger scale factor and, as a result, the selection of class of technique is less important. On the contrast, the selection of class of technique is very important if the scale factor is low. To demonstrate the superiority of the technique comparing with others, if the LR image is generated by low-pass filtering, a small-scale factor is recommended.

However, for the LR images obtained by DWT with db. 9/7, the result of sensitivity analysis is different, as illustrated in Figure 4-4 (d), (e) and (f). The values of std show that the selection of class of technique has significant effect on the results, and it is almost independent on the scale factor. In other words, the selection of scale factor to demonstrate the superior of a new technique is not important. Another observation is that the above conclusions are almost independent on test images due to the fact that Figure 4-4 (a), (b) and (c) have similar patterns, as well as Figure 4-4 (d), (e) and (f).

To demonstrate the visual quality of the produced results, Elaine image was selected from the tested images. Figure 4-5 shows the resolution enhanced images using the conventional interpolation methods and state-of-the-art resolution enhancement methods with an enlargement factor 4. The original HR image with the size of 512×512 pixels which is considered as the reference image to evaluate the performance of the implemented methods. Based on the image observation model, the input LR images with the size of 128×128 pixels were generated from blurring and down-sampling the original HR image through applying twice cascaded DWT with db.9/7 wavelet function. In spite of the impressive quantitative performance, the visual results show that the difference between the resolution enhanced images from the considered methods can be small, and it is difficult to be inspected visually. Note that all methods including, the proposed method and other selected methods, were implemented using Matlab 2015.

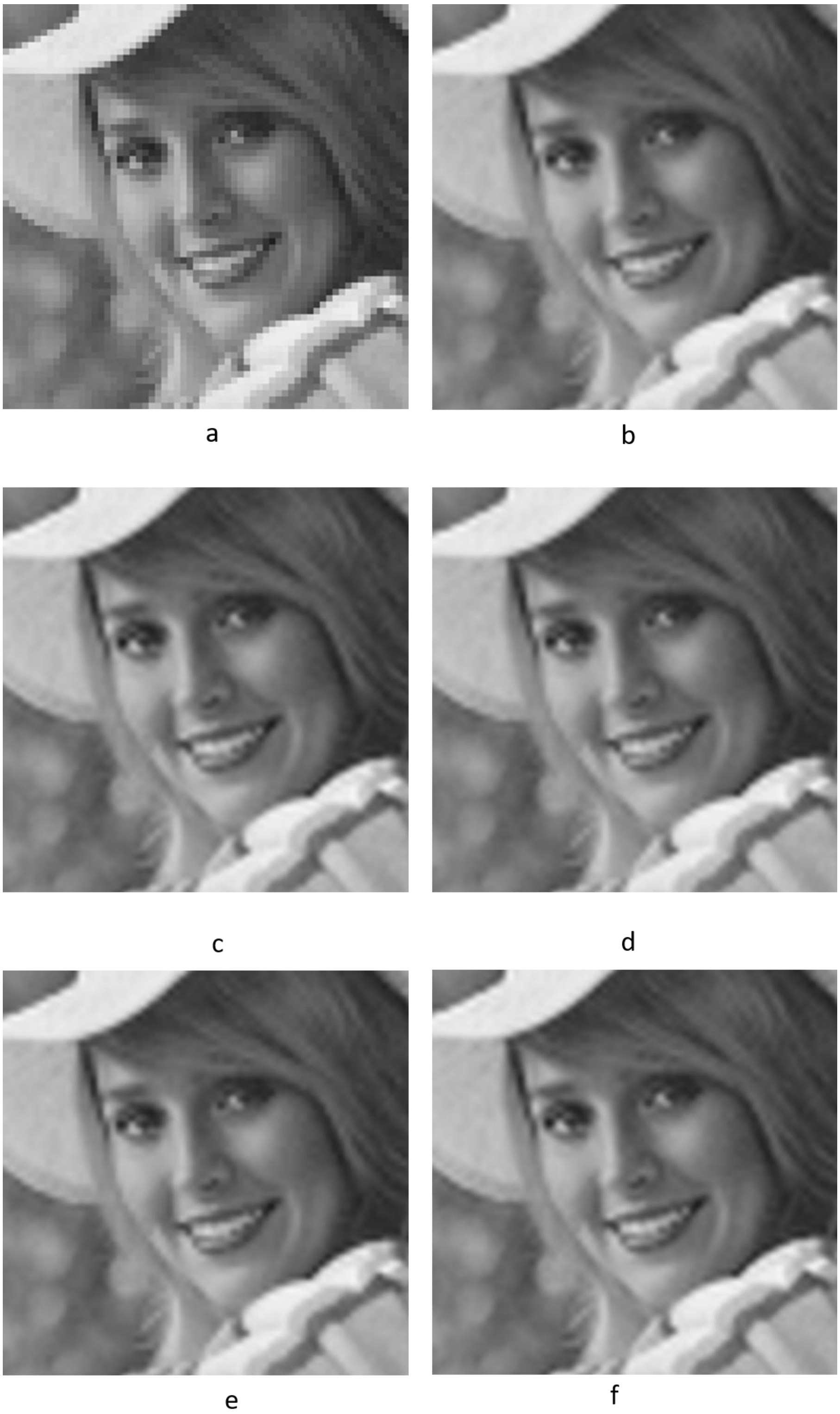


Figure 4-5 Resolution enhancement results from different interpolation methods and wavelet-based resolution enhancement methods with an enlargement factor from 128×128 to 512×512 for Elaine image (a) the input LR image, (b) Bicubic, (c) Lanczos, (d) WZP-CS, (e) WZP(db.9/7), and (f) WZP(coif2).

4.5 Summary

The wavelet-based image resolution enhancement techniques have been reviewed in this chapter, especially the way to assess the performance. The inconsistency of assumptions has been observed, and for some methods, the description of these assumptions is either neglected or unclear. Due to the fact that the laws of physics to generate LR images are unclear and also various case by case, the current methods to assess the performance may result in a biased conclusion. This chapter initially identifies the factors that substantially affect the performance of these techniques and quantitatively evaluates the impact of the existing assumptions. The importance of each factor has then been analysed by varying this factor and fixing the other factors. It has been revealed that the way of producing LR image, the variation of wavelet family and its wavelet functions, and the scale factor can significantly affect the performance of techniques. It has been also revealed that the selection of test images with different features as well as the selectin of interpolation method can influence moderately on the performance.

An Optimal Factor Analysis OFA approach has been proposed in this chapter in order to improve the performance of existing techniques and better evaluate the overall performance. The OFA algorithm selects the optimal technique (including the selection of wavelet family as well as its wavelet function and interpolation method) by simultaneously varying the way of producing LR image, the enlargement factor, and the test image. Moreover, a new Figure of Merit FOM measure is proposed to assist the selection of factors and better assess the overall performance.

The quantitative results reveal that the proposed OFA method can significantly improve the performance of the WZP methods. It also has the potential to be extended to other wavelet-based methods. Results also reveal that the most important factors that have significant effect on the performance are the method of producing LR image and the selection of wavelet function. For most of existing wavelet-based resolution enhancement techniques, the variation of these factors is very limited or never considered. The experimental results also indicate that the interpolation method has no significant effect on the performance and the best interpolation method is not consistent for different techniques. More precisely, the selection of interpolation method for wavelet-based techniques can affect the performance, but this effect is not distinct. For the LR images obtained by down-sampling using DWT with db.9/7, nearest neighbour, and low-pass filtering, wavelet-based techniques have the biggest potential to overtake the conventional interpolation methods. However, for the LR images produced by DWT with Haar, Bicubic interpolation, and Bilinear interpolation, wavelet-based techniques have no pronounced improvements over conventional interpolation methods. All these observations conclude that in order to assess more comprehensively and equitably for wavelet-based resolution enhancement techniques, variation of LR image generation method, scale factor, and wavelet functions must be considered; otherwise observed performance could be limited and inaccurate.

The limitation of the proposed OFA method is that it has relatively high computational cost and time consuming because this algorithm depends on the exhausted search to find the best parameters. Therefore, there is a need to develop a more computational efficient method, which has low computational cost and is suitable for real applications. The linkage from Chapter 4 to Chapter 5 is the contribution of investigation the important factors that affect the performance of the wavelet-based methods. This contribution plays a key role for improving the performance and presents a direct impact to connect between these two Chapters, e.g. it has been observed in Chapter 4 that the selection of wavelet function plays an important role in improving the performance. Therefore, the sensitivity analysis of wavelet function is conducted in Chapter 5 to fully explore the potential of the proposed method.

5 Satellite Image Resolution Enhancement using Discrete Wavelet Transform and New Edge-Directed Interpolation

5.1 Introduction

Satellite imaging is usually the most cost-efficient means of collecting regular and frequent data about the earth's surface. These data are routinely used to monitor land use change, urban expansion, agricultural health and productivity, the status of urban tree corridors, fire threat, environmental condition, etc. [136]. High-resolution (HR) satellite images are strongly demanded in many applications, such as remote sensing, astronomy, geoscience, and geographical information systems, not only for providing better visualisation but also for extracting additional information details, which can be crucial in these applications. For example, HR satellite images are crucial in distinguishing an object from similar ones and achieving a better classification of regions in a multi-spectral remote sensing image [1].

When remote sensing image data are acquired by sensors on satellites and aircraft, the data can have errors in geometry and in the measured intensity values of the pixels. The first is referred to as geometric errors while the latter is referred to as radiometric errors [137], [138]. The geometric errors can result from many sources, such as the relative motions of the remote sensing platform, nonidealities in the sensors themselves, the rotation of the earth, and uncontrolled variations in the position and attitude of the platform. However, the radiometric errors can arise from the instruments used to acquire the image data, the wavelength dependence of solar radiation, and the effect of the atmosphere. Geometric distortion effects of the image data are more severe than radiometric distortion and lead to varying degrees of severity in the produced image. Therefore, image processing procedures are used to compensate for these errors and to find more general applications. Resolution enhancement is one of the applications for correcting the errors in image geometry [137]. Resolution enhancement based on a single low-resolution (LR) image or multiple LR images, also called super-resolution (SR), recently has attracted lots of interest and has been used for different applications, such as satellite imaging [3, 66, 67, 71], medical imaging [5, 73], and video surveillance systems [7].

Interpolation is one of the commonly used techniques for image resolution enhancement. Fundamentally, it is the process of estimating values at unknown locations using known data [19]. There are four well-known conventional interpolation methods: nearest neighbour, bilinear, bicubic, and Lanczos. However, these linear methods cannot handle the fast-growing statistics around edges and accordingly yield interpolated images with blurred edges and undesirable artefacts. To address this problem, other nonlinear interpolation-based resolution enhancement methods have been developed to improve the subjective quality by taking edge information into account, such as edge-directed interpolation (EDI) [86] and new edge-directed interpolation (NEDI) [87]. EDI emphasises the visual integrity of the edges, and NEDI is the upgraded version. Nevertheless, the improvements by these methods are limited at the textures and non-linear edges of the interpolated images [78]. However, a lot of research has achieved enhanced performance of NEDI because the NEDI method uses a relatively simple model and hence has low computational complexity. The modified NEDI method was presented in [88] by considering a modified training widow structure to eliminate the prediction error accumulation and extending the covariance values into multiple directions to mitigate the covariance mismatch problem.

Another class of image resolution enhancement methods is wavelet-based. A common assumption of wavelet-based methods is that the LR image is the low-pass filtered sub-band of the wavelet-transformed HR image. Estimating the unknown wavelet coefficients in sub-bands containing high-pass frequency spatial information is the essential target of this class of algorithms to estimate the HR image from the LR image. Many advanced methods have been introduced in order to estimate the wavelet coefficients of high-frequency sub-bands.

This chapter proposes an improved image resolution enhancement approach based on discrete wavelet transform (DWT) and new edge-directed interpolation (NEDI) for degraded satellite images by geometric distortion to correct the errors in image geometry and recover the edge details of directional high-frequency sub-bands, called DWT-NEDI, which integrates merits from both the frequency-domain and spatial-domain. There are three major stages in the proposed method. First, the DWT is employed to decompose the input image into different frequency sub-bands in the frequency-domain. Second, for the three high-frequency sub-bands, NEDI is employed to process each of them and output the estimated sub-bands using soft-thresholding method. Meanwhile, the input image is directly interpolated by NEDI to output the low-frequency sub-band. Finally, after combining both low and high-frequency sub-bands, the processed image is transformed back to the spatial-domain as the outcome of resolution enhancement through inverse discrete wavelet transform (IDWT). The majority of developed SR approaches focus on grayscale or single-channel image SR, while this study applies the performance of the proposed technique to colour images. Although the proposed approach also works well for other types of images, this chapter focuses its application on satellite images.

5.2 State-of-The-Art Image Resolution Enhancement Methods

In this section, the state-of-the-art new edge-directed interpolation (NEDI) and state-of-the-art wavelet-based image resolution methods, including discrete wavelet transform (DWT)-based resolution enhancement methods will be described below as examples of the state-of-the-art resolution enhancement methods.

5.2.1 New Edge-Directed Interpolation (NEDI)

Li and Orchard [87] proposed a new edge-directed interpolation NEDI method for upsampling regular sampled images. The fundamental idea is to estimate local covariance coefficients from a LR image and then employ the covariance estimates to adapt the interpolation coefficients by exploiting the geometric duality between the LR covariance and the HR covariance. Geometric duality refers to the matching between the HR covariance and the LR covariance which couple the pair of pixels along the same orientation but at the different resolution.

Let consider the LR image $Y_{i,j}$ with size of $H \times W$ obtained from the original HR image $X_{2i,2j}$ with size of $2H \times 2W$, which means $X_{2i,2j} = Y_{i,j}$. To interpolate the interlacing lattice (the missing pixel) $X_{2i+1,2j+1}$ from the lattice $X_{2i,2j}$, the forth-order linear interpolation can be used, as shown in

$$\hat{X}_{2i+1,2j+1} = \sum_{k=0}^{1}\sum_{l=0}^{1} \propto_{2k+l} X_{2(i+k),2(j+l)} \tag{5-1}$$

The interpolation is performed by weighted averaging the four nearest neighbouring pixels along the diagonal directions. Using the classical Wiener filtering theory [139], the optimal minimum mean squared error (MMSE) linear interpolation coefficients can be given by

$$\vec{\propto} = R^{-1}\vec{r} \tag{5-2}$$

where $R = [R_{k,l}]$, $(0 \le k, l \le 3)$ and $\vec{r} = [r_k]$, $(0 \le k \le 3)$ are the HR covariance. The computation of HR covariance $R_{k,l}, r_k$ require the knowledge of $X_{2i+1,2j+1}$ which is not available before interpolation. To overcome this difficulty, the geometric duality is used. The geometric duality between the HR covariance $R_{k,l}, r_k$ and the LR covariance $\hat{R}_{k,l}, \hat{r}_k$ is shown in Figure 5-1, where the LR covariance $\hat{R}_{k,l}, \hat{r}_k$ is applied to replace the HR covariance $R_{k,l}, r_k$. Therefore, the unknown pixel (the missing pixel) $X_{2i+1,2j+1}$ is estimated by Eq. (5-1) with $\hat{R}_{k,l}, \hat{r}_k$ [88].

The advantage of NEDI algorithm is that it substantially improves the subjective quality of the pixels around edges. In order to perform a better trade-off between subjective quality and computational complexity, a hybrid approach of combining bilinear interpolation and covariance-based-adaptive interpolation is achieved. Covariance-based adaptive interpolation is only applied to pixels around edges (edge pixels) while bilinear interpolation is only applied to pixels in the smooth regions (non-edge pixels).

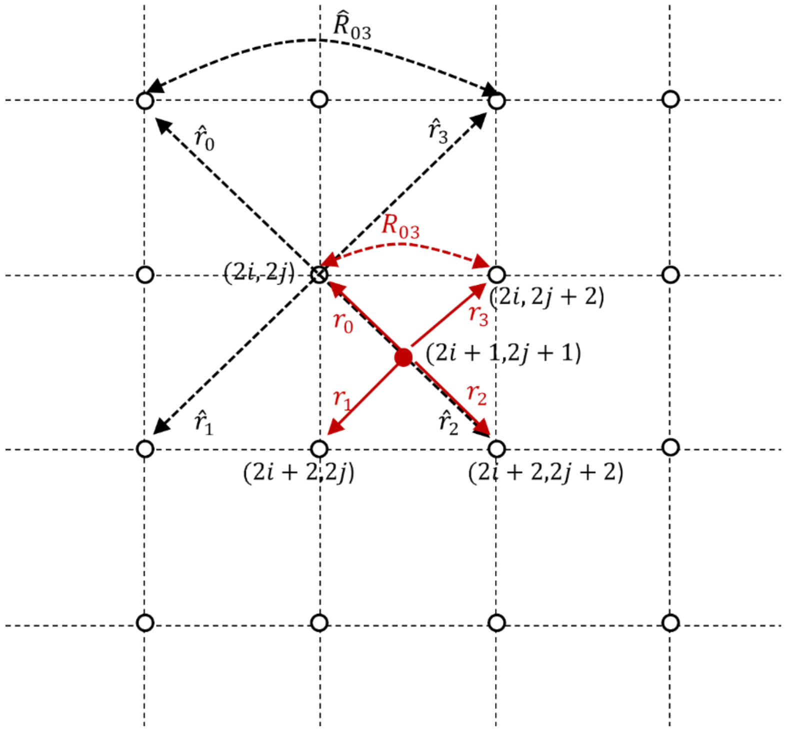


Figure 5-1 Geometric duality between the HR covariance $R_{k,l}, r_k$ and the LR covariance $\hat{R}_{k,l}, \hat{r}_k$ is exploited when interpolating the missing pixel $\hat{X}_{2i+1,2j+1}$ from $X_{2i,2j}$ using NEDI

5.2.2 DWT-based Image Resolution Enhancement Methods

An edge-enhanced DWT-based image up-sampling method was proposed by Acharya and Tsai [69]. One-level DWT process is employed to decompose an input LR image into four frequency sub-bands (LL, LH, HL, and HH). After the decomposition, each sub-band has quarter the size of the input image due to down-sampling operation in each of the DWT sub-bands. Then the low-frequency LL sub-band is replaced by the input image as well as multiplied by the scale factor α, and the high-frequency sub-bands LH and HL are up-sampled with the factor of 2 by inserting zeros between successive rows and columns in each sub-band and HH is discarded. IDWT process is performed on these approximated sub-bands to generate an up-sampled image. This method was demonstrated better visual quality than the conventional interpolation methods. Figure 5-2 shows a block diagram of the edge-enhanced DWT-based up-sampling technique.

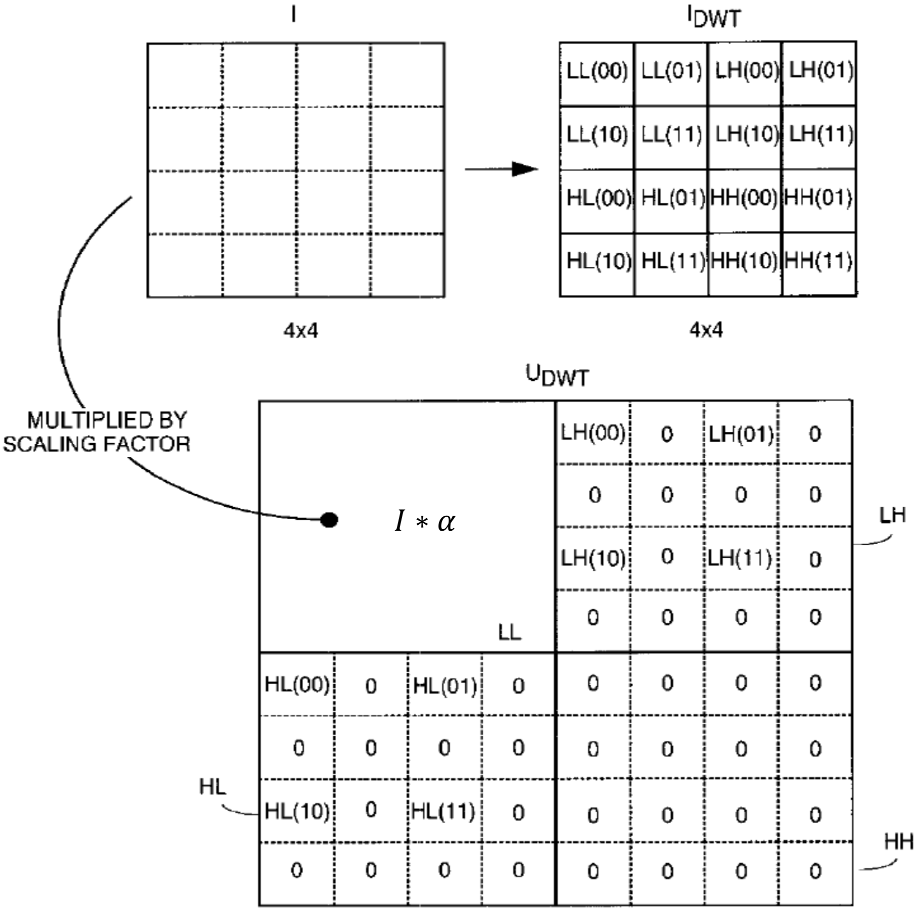


Figure 5-2 Block diagram of the DWT-based up-sampling method [69]

A Demirel-Anbarjafari Super-Resolution (DASR) method [70] was introduced based on the interpolation of the input image and the high-frequency sub-bands obtained through the DWT process. In the DASR algorithm, the three high-frequency sub-bands are interpolated using bicubic interpolation with the enlargement factor α. The LL sub-band is replaced by the input image because it contains more information and is also interpolated using bicubic interpolation with half of the scale factor α/2 used to interpolate the high-frequency sub-bands in order to have the required size for the IDWT process. Finally, IDWT is applied to achieve a resolution enhanced image by combining the interpolated sub-band images and the input image. Although the DWT process was employed to preserve the high-frequency components in the high-frequency sub-bands of the image, the blurring effect from the employed interpolation method causes the potential loss of edges in these sub-bands.

5.2.3 DWT and Difference Image (DWT-Diff)-based Image Resolution Enhancement Methods

Demirel and Anbarjafari proposed a DWT-Diff technique [71] for resolution enhancement of satellite images based on interpolating the input image as well as the high-frequency sub-band images, and correcting the estimated high-frequency components using the difference image. In the DWT-Diff algorithm, all the decomposed frequency sub-bands are interpolated using bicubic interpolation with factor 2. To preserve more edges, an intermediate stage using the difference image is included for enhancing the estimated high-frequency components by subtracting the interpolated LL sub-band from the input LR image and then adding the difference image into the estimated high-frequency sub-bands, as illustrated in Figure 5-3. Another interpolation by using bicubic interpolation with factor $\alpha/2$ can be applied again to the new corrected sub-bands to have the required size for the IDWT process. In parallel, the input image is also interpolated with factor $\alpha/2$. A sharper enhanced image is obtained by applying IDWT to combine all these sub-band images.

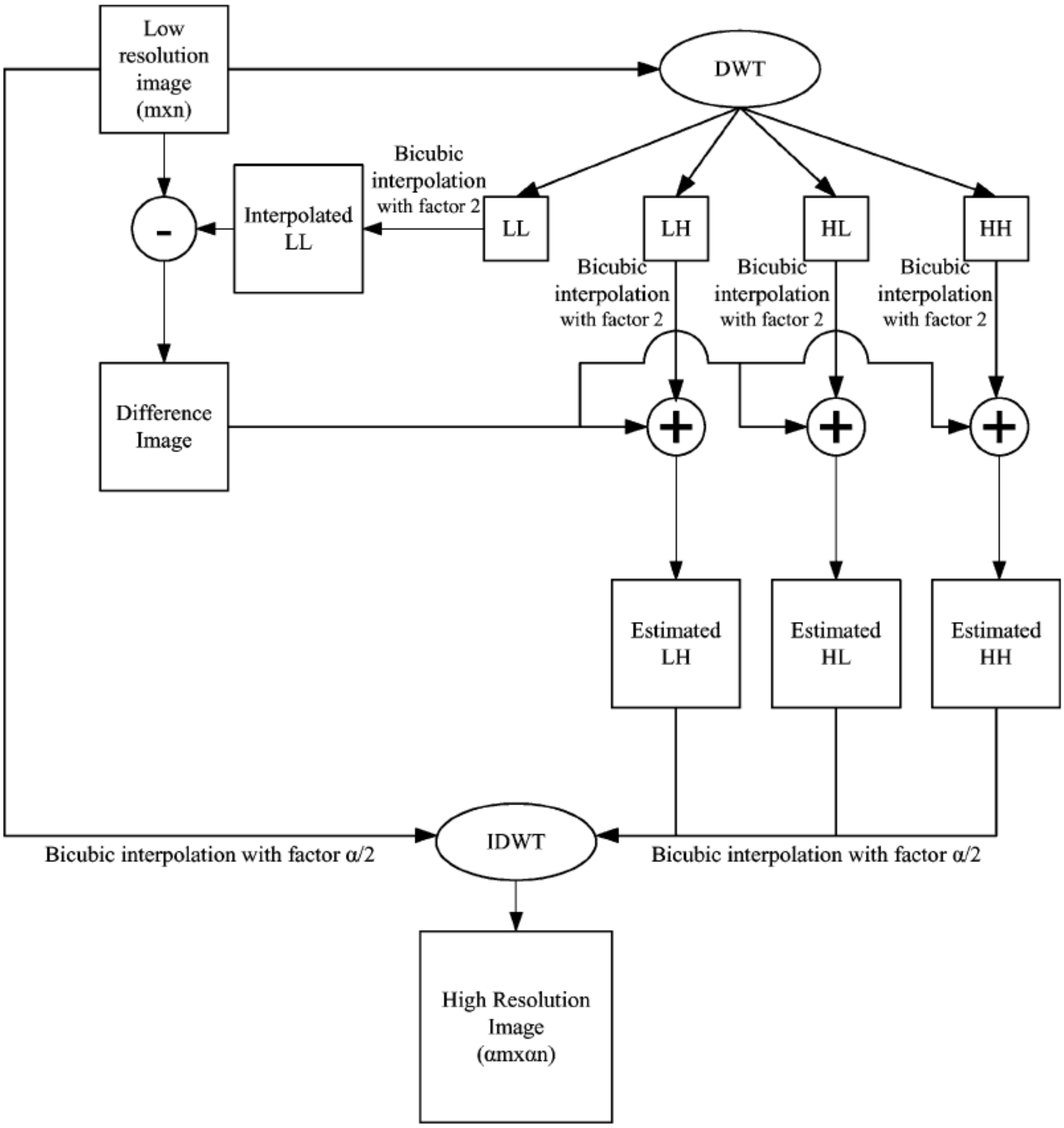


Figure 5-3 Block diagram of the DWT-Diff technique [71]

5.2.4 DWT and Stationary Wavelet Transform (DWT-SWT)-based Image Resolution Enhancement Methods

A DWT-stationary wavelet transform SWT technique [72] was proposed based on interpolating the high-frequency sub-bands as well as the input image, and correcting the estimated high-frequency sub-bands using SWT. One-level DWT is used to separate an input LR image to obtain high-frequency sub-bands, and then these sub-bands are interpolated using bicubic interpolation with factor 2. To better preserve the edge details, an intermediate process using SWT is included for correcting the estimated high-frequency sub-bands by adding the high-frequency sub-bands obtained through SWT of the input image with the interpolated DWT high-frequency sub-bands. For higher enlargement, bicubic interpolation by the factor of α/2 can be applied further to the new corrected high-frequency sub-bands. The input image is used for interpolation instead of using the LL sub-band and is interpolated with the factor of α/2. Finally, all these sub-bands are combined using IDWT to achieve a sharper HR image. The block diagram of this algorithm is shown in Figure 5-4.

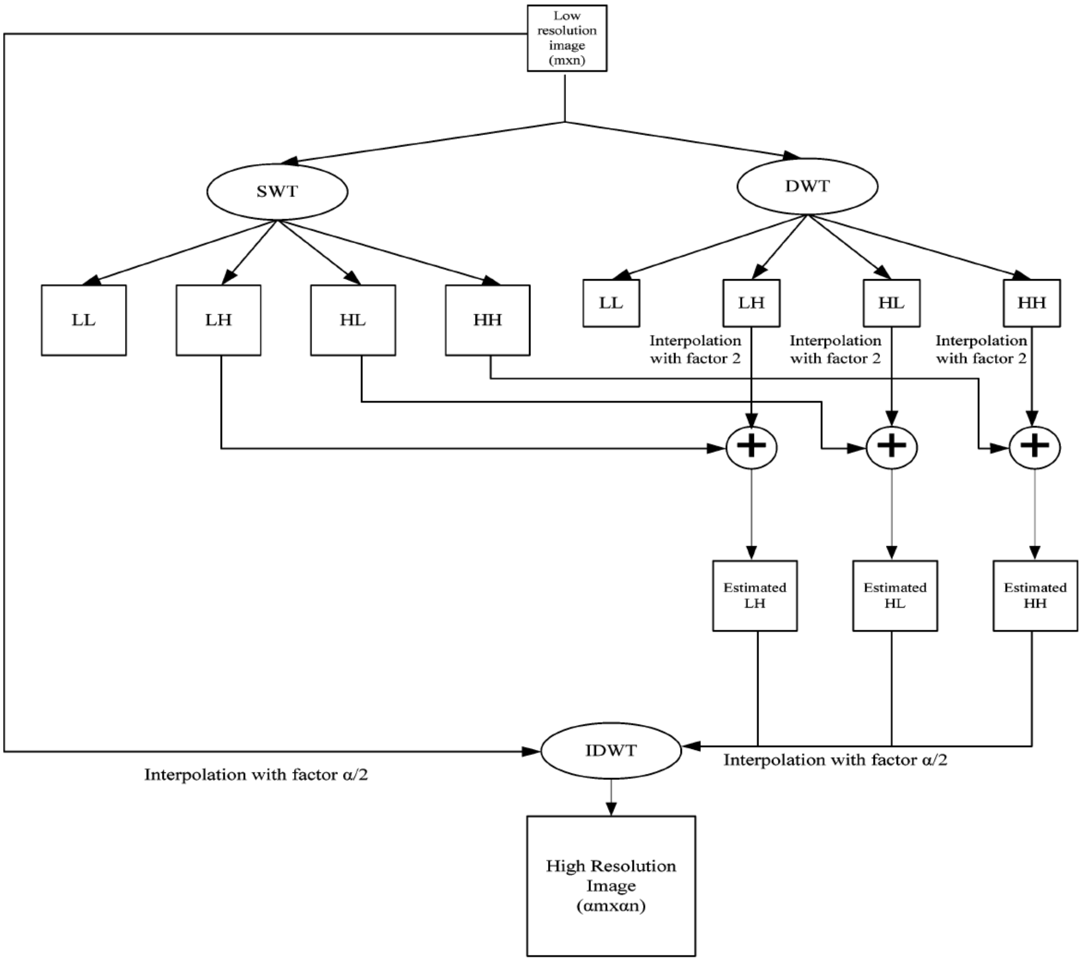


Figure 5-4 Block diagram of the DWT-SWT technique [72]

5.3 Proposed Resolution Enhancement Approach

The main drawback for using a direct interpolation to enhance image resolution is the generated visual degradations around edge areas and, as a result, the production of blurred edges. This degradation is due to annoying levels of smoothing across edges caused by the employed interpolation method, which does not use any information pertinent to the edges in the original image. Preserving the high-frequency components (i.e. edges) and consequently increasing the quality of the resulted HR image is the fundamental target for reconstructing the HR image from the given LR image.

Therefore, in this work, the DWT process was employed to isolate and better preserve the edges of the image using the interpolation of high-frequency sub-bands. This is due to the interpolation of isolated high-frequency components in the high-frequency sub-bands preserving more edges of the image than interpolating the image directly. On the other hand, the DWT process decomposes the given image into three directional high-frequency sub-bands, which isolate the edges in three directions and thus reduce the undesirable inter-directional interference in the resolution enhancement process. Although a number of DWT-based interpolation methods [69, 70, 71, 72] have been developed for recovering the missing high-frequency components of the given LR image, the blurring effect from their employed interpolation methods causes the potential loss of edges in these sub-bands. For example, bicubic, the most widely used interpolation method in wavelet-based resolution enhancement approaches, can produce blurring around edge areas because of the smoothing process. Therefore, the blurring effect caused by the interpolation method needs to be addressed.

Many dual-tree complex wavelet transform (DT-CWT)-based resolution enhancement methods [68, 74, 77] attempt to address this problem. In [68], the edge-directed interpolation EDI [86] was employed as an alternative interpolation method of high-frequency sub-bands obtained by DT-CWT. Later, Izadpanahi and Demirel [74] proposed an extended version of this approach for multi-frame SR. Recently, the same authors applied a new-edge directional interpolation NEDI [87] which improved the performance of EDI, for the interpolation of high-frequency sub-bands generated by DT-CWT for motion-based video SR [77]. In this work, NEDI is employed to process the high-frequency sub-bands obtained through DWT and output the estimated sub-bands using a nonlinear adaptive threshold.

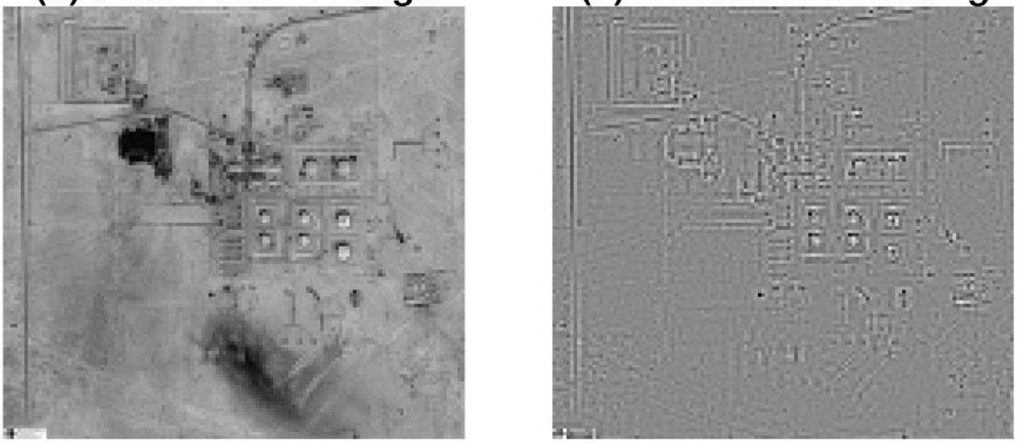
This chapter proposes combining DWT and NEDI, which integrates merits from both the frequency-domain and spatial-domain, and substantially improves the visual quality of the pixels around edges. The advantage of using DWT with NEDI is that it recovers the edge details of directional high-frequency sub-bands and thus decreases the annoying inter-directional interference in the SR process. This merit cannot be achieved using only the NEDI method. Consider an input LR image with the size of $W \times H$, and the scale factor is denoted by α. Initially, one-level DWT process decomposes the input LR image into four frequency sub-bands, called low-low (LL), low-high (LH), high-low (HL), and high-high (HH). Each sub-band has half the size of the input image due to down-sampling. The high-frequency sub-bands (LH, HL, and HH) are interpolated using the NEDI method with the scale factor α. Generally, image data sets recorded by sensors are most commonly corrupted by additive Gaussian noise and multiplicative noise. Therefore, to preserve more edges and reduce the noise in the estimated high-frequency sub-bands, a thresholding procedure using an adaptive threshold is included to process the produced wavelet coefficients. Many types of thresholding functions have been introduced for the modification of estimated wavelet coefficients, such as hard and soft [127]. This chapter employs a nonlinear soft-thresholding technique proposed by Donoho [140] and extended by Zhang [141]. The universal threshold τ for the considerd sub-band can be calculated as

$$\tau = \sigma\sqrt{2\log(N)\,/N} \tag{5-3}$$

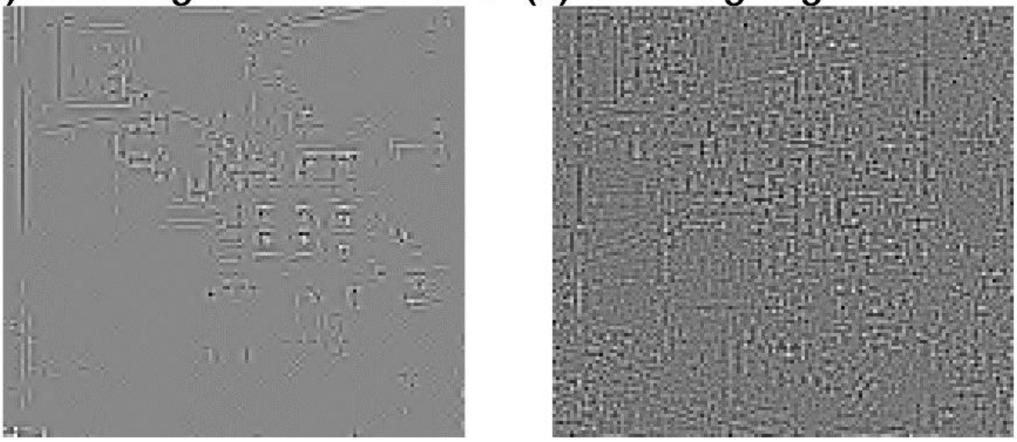
where σ is the standard deviation of the sub-band and N is the total number of pixels. The nonlinear soft-thresholding function is defined as

$$X_{out}(i,j) = \begin{cases} X_{in}(i,j) - \tau & X_{in}(i,j) > \tau \\ 0 & |X_{in}(i,j)| \le \tau \\ X_{in}(i,j) + \tau & X_{in}(i,j) < -\tau \end{cases} \tag{5-4}$$

The basic idea of this thresholding process is that the energy of a signal is often concentrated on a few coefficients while the energy of noise is spread among all coefficients in the wavelet domain. Therefore, the nonlinear soft-thresholding tends to maintain few larger coefficients representing the signal while reducing noise coefficients to zero in the wavelet domain. A universal threshold is intuitively expected to uniformly remove the noise since the Gaussian noise still has the same variance over different scales in the transform domain [141]. The application of this soft-thresholding function is based on the hypothesis that the large coefficients in the high-frequency sub-bands reflect the true edges of objects while the small coefficients reflect the noise. This hypothesis can be proven by Figure 5-5, where (a) shows the observed LR image and (b) shows the reconstruction image of high-frequency sub-bands only without thresholding. Both true edges and noise can be observed in Figure 5-5(b). Figure 5-5(c) shows the reconstruction image of high-frequency sub-bands with thresholding where the small coefficients are removed. It can be observed that the noise is significantly reduced while the true edges are preserved. The reconstruction image of high-frequency sub-bands where the large coefficients are removed is illustrated by Figure 5-5(d), which is dominated by noise with very little true edge information found.



(a) observed LR image (b) without thresholding



(c) removing small coeffcients (d) removing large coeffcients

Figure 5-5 An example to help justify the use of thresholding process. (a) the observed LR image; (b) the reconstruction image of high-frequency sub-bands only without thresholding; (c) the reconstruction image of high-frequency sub-bands only where the small coefficients are removed; (d) the reconstruction image of high-frequency sub-bands only where the large coefficients are removed.

The input LR image, interpolated by the NEDI method with half of the scale factor α/2, is used as the estimated LL sub-band because it contains more information than the LL sub-band produced by the DWT process, as suggested in [70, 71, 72]. Finally, IDWT is applied to achieve a super-resolved image by combining the estimated LL sub-band and corrected high-frequency sub-bands. The block diagram of the DWT-NEDI approach is illustrated by Figure 5-6.

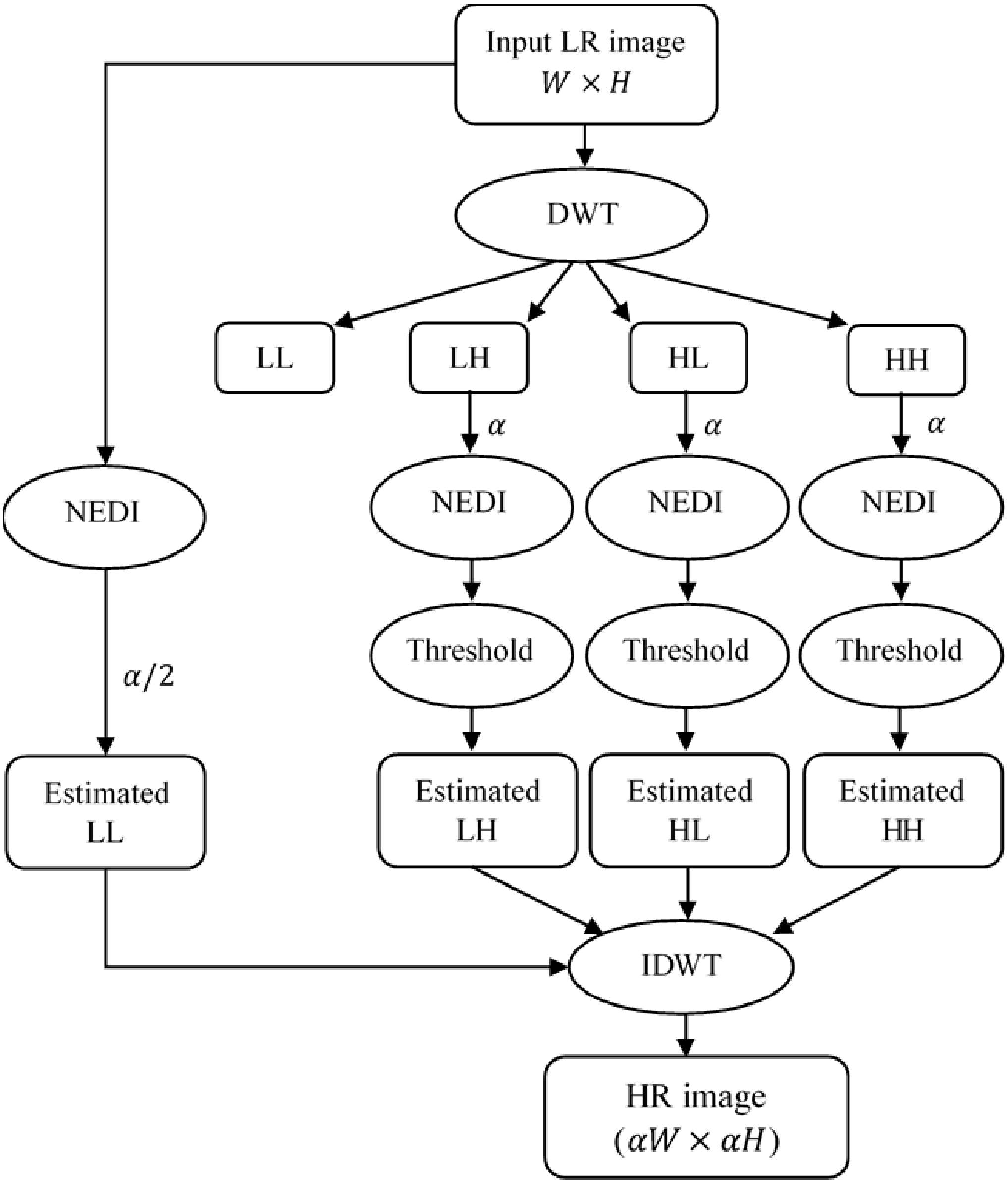


Figure 5-6 Block diagram of the proposed DWT-NEDI resolution enhancement approach.

The proposed approach can be summarised by the following steps:

1. Consider the red channel of the observed LR image;
2. Compute one-level DWT decomposition of this channel;

3. Apply the NEDI method to LH, HL, and HH high-frequency sub-bands with the scale factor α;
4. For each high-frequency sub-band, calculate the threshold τ;
5. Apply the adaptive thresholding process for each high-frequency sub-band and create the estimated $\widehat{\mathrm{LH}}$, $\widehat{\mathrm{HL}}$ and $\widehat{\mathrm{HH}}$;
6. Apply the NEDI method to the input LR image with the scale factor $\alpha/2$ to create $\widehat{\mathrm{LL}}$;
7. Apply IDWT using $(\widehat{\mathrm{LL}}, \widehat{\mathrm{LH}}, \widehat{\mathrm{HL}}, \widehat{\mathrm{HH}})$ to produce the enhanced channel;
8. Repeat steps 2-7 for the blue and green channels respectively;
9. Combine the three enhanced channels into the final enhanced HR colour image.

5.4 Results and Discussions

5.4.1 Visual and Quantitative Evaluation of Standard Images

This section presents the experimental results of visual and quantitative evaluation to firstly test the proposed DWT-NEDI method on the same standard images used in Chapter 4 before testing on the more challenge satellite images. The proposed method was tested on three well-known standard images "*Lena, Baboon, and Elaine*". The original test HR images with size of 512×512 pixels were considered as the reference images to evaluate the performance of the proposed algorithm. Based on the observation model, the input LR images with the size of 128×128 pixels were generated from blurring and down-sampling the original HR images through achieving twice cascaded DWT by the db.9/7 wavelet function and considering the low-frequency (LL) sub-band of DWT. The observed LR images were further corrupted by an additive Gaussian noise with the signal-to-noise ratio (SNR) of $40\ \mathrm{dB}$.

To illustrate the visual quality of the experimental results, Elaine image was selected because of the most improvement with the proposed DWT-NEDI algorithm. Figure 5-7 shows the resolution enhanced images using the proposed method and the other compared methods with an enlargement from 128×128 pixels to 512×512 pixels of the Elaine image. It can be observed that the DWT-NEDI algorithm preserves the sharp edges of the original image better than the other methods.

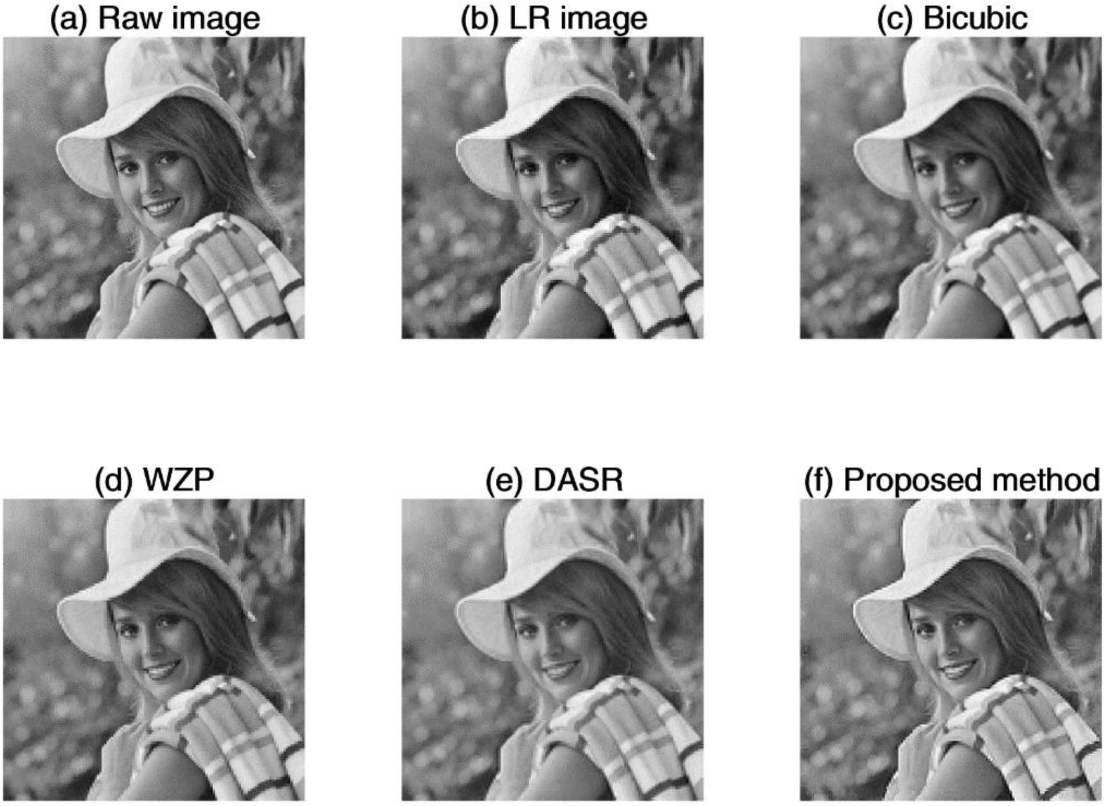


Figure 5-7 Resolution enhanced results from the proposed approach with an enlargement from 128×128 to 512×512 for an image selected from the *construction* group. (a) the whole input LR image; (b) the selected region of the input LR image; resolution enhance images by (c) Bicubic, (d) WZP, (e) DASR and (f) the proposed method.

Tables 5-1 to 5-4 respectively, list the quantitative results evaluated by PSNR, RMSE, Entropy, and SSIM for all three tested standard images. In terms of the values of PSNR and RMSE, the proposed DWT-NEDI method has the greatest improvement among the considered methods for Elaine image with many sharp edges and Lena image with many strong edges relative to the number of textures, while the WZP-CS method has the highest improvement for Baboon image. In terms of the results of Entropy, the DWT-NEDI technique has the best performance for Elaine image with edges and Baboon image with relative abundance of textures, whereas the WZP method is the best for Lena image. In terms of the values of SSIM, the WZP-CS method produces the highest performance for all three tested images. Nevertheless, the method proposed still has a higher improvement better than the other remaining methods. These observations conclude that the performance of the technique proposed depends on different image features and performance evaluation criteria.

Table 5-1 PSNR (db) results of the selected standard images for resolution enhancement from 128×128 to 512×512.

Techniques	Standard Images		
	Lena	Baboon	Elaine
Nearest	23.76	19.60	25.65
Bilinear	24.92	20.06	26.88
Bicubic	24.84	20.03	26.88
Lanczos	24.75	20.00	26.83
WZP	25.50	20.21	27.48
WZP-CS	26.34	**20.70**	28.52
DWT	25.40	20.16	27.38
DASR	25.19	19.88	27.60
DWT-Dif	23.52	19.29	25.41
DWT-SWT	24.42	19.53	26.69
Proposed	**26.91**	20.31	**29.12**

Table 5-2 RMSE results of the selected images for resolution enhancement from 128×128 to 512×512.

Techniques	Standard Images		
	Lena	Baboon	Elaine
Nearest	16.54	26.71	13.31
Bilinear	14.48	25.33	11.55
Bicubic	14.60	25.41	11.55
Lanczos	14.76	25.50	11.61
WZP	13.53	24.89	10.79
WZP-CS	12.29	**23.53**	9.56
DWT	13.69	25.04	10.91

DASR	14.03	25.85	10.64
DWT-Dif	17.01	27.66	13.67
DWT-SWT	15.33	26.91	11.81
Proposed	**11.51**	24.61	**8.92**

Table 5-3 Entropy results of the selected images for resolution enhancement from 128×128 to 512×512.

Techniques	Standard Images		
	Lena	Baboon	Elaine
Nearest	4.97	6.20	4.93
Bilinear	5.38	6.14	5.39
Bicubic	5.29	6.14	5.29
Lanczos	5.32	6.15	5.28
WZP	**4.86**	6.09	4.91
WZP-CS	5.20	6.41	5.28
DWT	4.88	6.10	4.92
DASR	5.15	6.18	5.35
DWT-Dif	5.00	6.14	5.36
DWT-SWT	5.17	6.08	5.28
Proposed	5.18	**6.03**	**4.58**

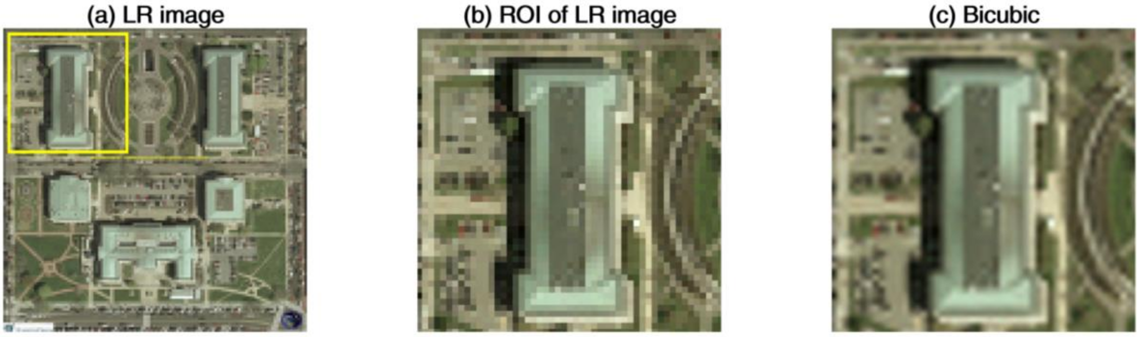
Table 5-4 SSIM results of the selected images for resolution enhancement from 128×128 to 512×512.

Techniques	Standard Images		
	Lena	Baboon	Elaine
Nearest	0.30	0.19	0.37
Bilinear	0.36	0.20	0.43
Bicubic	0.37	0.22	0.44
Lanczos	0.37	0.23	0.44
WZP	0.39	0.27	0.43
WZP-CS	**0.43**	**0.29**	**0.48**
DWT	0.38	0.27	0.43
DASR	0.36	0.24	0.42
DWT-Dif	0.29	0.21	0.34
DWT-SWT	0.33	0.22	0.39
Proposed	0.40	0.27	0.45

5.4.2 Visual Evaluation of Satellite Images

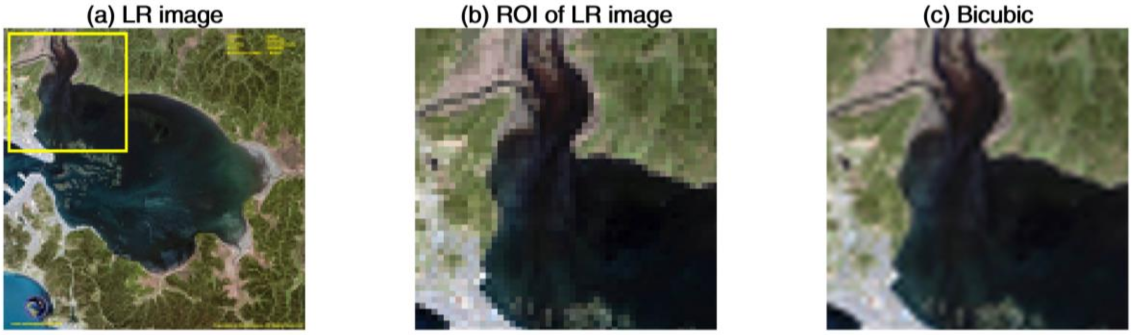
The reason to choose the more challenging satellite images as test images to improve the spatial resolution is that satellite imaging is one such an application as improving the sensors on satellites is hugely expensive. In addition, analysis from these low-quality (degraded) images can be extremely difficult. Furthermore, SR is still highly demanded in satellite imaging although the hardware specification has been improved. This is because it is always attractive to see smaller objects with more details. The proposed technique was tested on 20 different satellite images obtained from the Satellite Imaging Corporation webpage [142], which is a public data set. To assess the variation in performance in terms of different types of images, the studied satellite images were divided into five classes: "*natural disaster*, *tourism*, *defence and intelligence*, *construction*, and *cadastre and land*". Each class includes four images that were randomly selected from the data set. The size of the original HR images in the public data set is different. For the consistency of comparison, each original HR image was therefore resized to 512×512 pixels as the reference image. The input LR images with the size of 128×128 pixels were produced from blurring and down-sampling the original HR images by applying twice cascaded DWT with the db.9/7 wavelet filter. The observed LR images were further corrupted by a Gaussian noise with the signal-to-noise ratio (SNR) of $40\,\mathrm{dB}$. The biorthogonal Daubechies (db.9/7) was chosen because the literature review shows that it is the most commonly used wavelet function for the decomposition process by DWT [132]. Note that all methods including, the proposed method and other considered methods, were implemented using Matlab 2015.

To demonstrate the visual quality of the produced results, one image was randomly selected from each group and tested by the proposed technique. Figures 5-8 and 5-9 show the resolution enhanced images using the proposed technique and the other considered methods with an enlargement from 128 × 128 to 512 × 512 of the images from the *construction and tourism* groups. The visual results demonstrate the ability of the DWT-NEDI technique to enhance the observed LR images by proving more sharp edges, potentially offering more details of interested objects.


(a) LR image
(b) ROI of LR image
(c) Bicubic


(d) WZP
(e) DASR
(f) Proposed method

Figure 5-8 Resolution enhanced results from the proposed approach with an enlargement from 128×128 to 512×512 for an image selected from the *construction* group. (a) the whole input LR image; (b) the selected region of the input LR image; resolution enhanced images by (c) Bicubic, (d) WZP, (e) DASR and (f) the proposed method.



(a) LR image (b) ROI of LR image (c) Bicubic

(d) WZP (e) DASR (f) Proposed method

Figure 5-9 Resolution enhanced results from the proposed approach with an enlargement from 128×128 to 512×512 for an image selected from the *tourism* group. (a) the whole input LR image; (b) the selected region of the input LR image; resolution enhanced images by (c) Bicubic, (d) WZP, (e) DASR and (f) the proposed method.

5.4.3 Quantitative Evaluation of Satellite Images

The difference between the super-resolved images from different techniques can be small, and it is difficult to be inspected visually. This section presents the results of quantitative comparison. The peak-signal-to-noise-ratio (PSNR) between the super-resolved image and the original HR image is one of the most commonly used objective fidelity criteria for evaluating image quality. It can be calculated as Eq. (4-8).

MSE represents the mean-square-error between the super-resolved image $\hat{X}(i,j)$ and the original HR image $X(i,j)$. It can be calculated as

$$MSE = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \left[\hat{X}(i,j) - X(i,j)\right]^2 \quad \textbf{(5-5)}$$

The root-mean-square error (RMSE) between these two images is also one of the commonly used quantitative measures [19], and it can be expressed as

$$RMSE = \sqrt{MSE} \tag{5-6}$$

Entropy is another quantitative measure used to assess image quality when the error images for different image resolution enhancement techniques are very close to each other and it is very difficult to make assessment. The entropy of a negative error image, denoted by E, can be calculated as

$$E = -\sum_{k=1}^{L} P\,(r_k) log_2 P(r_k) \tag{5-7}$$

where $P(r_k)$ is the probability of an intensity value r_k. The lower the E, the better the improvement is [135].

To complement the quantitative analysis, the structural similarity (SSIM) [143] image quality measure has also been applied. The SSIM index evaluates the visual effect of three characteristics of an image: luminance, contrast, and structure. It is based on the computation of these three components and is an inner product of them. It is defined as

$$SSIM = \frac{(2\mu_{\hat{X}}\mu_X + C_1)(2\sigma_{\hat{X}}\sigma_X + C_2)}{\left(\mu_{\hat{X}}^2 + \mu_X^2 + C_1\right)\left(\sigma_{\hat{X}}^2 + \sigma_X^2 + C_2\right)} \tag{5-8}$$

where $\mu_{\hat{X}}, \mu_X$ are the local means for the images $\hat{X}, X$, respectively, $\sigma_{\hat{X}}$,σ_X are corresponding standard deviations, and C_1, C_2 are two constants used to avoid the instability. The higher the SSIM, the better the improvement is.

The quantitative performances measured by PSNR, RMSE, Entropy, and SSIM for the selected examples of each group are listed in Tables 5-5 to 5-8, respectively. In terms of the results of PSNR and RMSE, the proposed technique has the best performance for all five selected images. In terms of the results of SSIM, the proposed technique and WZP-CS achieve the highest performance for the images from the *natural disaster* and *tourism* groups, whereas the proposed technique is the best for the images from the remaining groups. In terms of the values of entropy, the nearest neighbour has the best performance for the images from the *natural disaster* and *tourism* groups, WZP and DWT-Diff produces the highest performance for the images from the *defence* and *cadastre* groups, while the proposed technique is the best for the images from the *construction* group.

Table 5-5 PSNR (db) results of the selected images for resolution enhancement from 128×128 to 512×512.

Techniques	Image Group*				
	1	2	3	4	5
Nearest	22.42	25.87	22.53	19.66	15.92
Bilinear	23.25	26.67	23.21	20.42	16.65
Bicubic	23.20	26.68	23.18	20.39	16.61
Lanczos	23.15	26.65	23.13	20.35	16.56
WZP	23.65	26.92	23.50	20.82	16.93
WZP-CS	24.17	27.57	23.79	21.27	17.34
DWT	23.54	26.83	23.42	20.71	16.85
DASR	22.99	26.48	22.90	20.00	16.51
DWT-Dif	21.92	25.47	21.89	18.94	15.57
DWT-SWT	22.31	25.87	22.37	19.43	15.99
Proposed	**24.90**	**27.83**	**24.53**	**21.93**	**18.01**

*Group 1: natural disaster; 2: tourism; 3: defence and intelligence, 4: construction, and 5: cadastre and land.

Table 5-6 RMSE results of the selected images for resolution enhancement from 128×128 to 512×512.

Techniques	Image Group				
	1	2	3	4	5
Nearest	19.31	12.98	19.06	26.51	40.80
Bilinear	17.55	11.84	17.62	24.30	37.50
Bicubic	17.63	11.83	17.68	24.39	37.68
Lanczos	17.74	11.87	17.78	24.51	37.91
WZP	16.75	11.50	17.04	23.20	36.30
WZP-CS	15.77	10.67	16.49	22.03	34.63
DWT	16.96	11.61	17.20	23.50	36.64
DASR	18.08	12.10	18.26	25.50	38.11
DWT-Dif	20.43	13.59	20.51	28.82	42.46
DWT-SWT	19.54	12.97	19.42	27.24	40.47

Proposed	**14.50**	**10.35**	**15.13**	**20.42**	**32.05**

Table 5-7 Entropy results of the selected images for resolution enhancement from 128×128 to 512×512.

Techniques	Image Group				
	1	2	3	4	5
Nearest	**5.01**	**4.83**	5.62	5.83	6.06
Bilinear	5.31	4.96	5.84	6.08	6.19
Bicubic	5.21	4.92	5.82	5.92	6.15
Lanczos	5.21	4.92	5.81	5.91	6.21
WZP	5.13	4.86	**5.60**	5.67	5.98
WZP-CS	5.20	4.87	5.79	5.65	6.11
DWT	5.04	4.89	5.62	5.59	5.98
DASR	5.17	4.85	5.70	5.87	6.13
DWT-Dif	5.20	4.86	5.71	5.71	**5.88**
DWT-SWT	5.17	5.02	5.66	5.84	5.90
Proposed	5.16	4.89	5.83	**5.54**	6.14

Table 5-8 SSIM results of the selected images for resolution enhancement from 128×128 to 512×512.

Techniques	Image Group				
	1	2	3	4	5
Nearest	0.27	0.26	0.21	0.24	0.17
Bilinear	0.31	0.29	0.21	0.26	0.19
Bicubic	0.33	0.31	0.23	0.28	0.21
Lanczos	0.33	0.31	0.24	0.28	0.21
WZP	0.34	0.34	0.29	0.33	0.25
WZP-CS	**0.38**	**0.37**	0.30	0.35	0.27
DWT	0.33	0.33	0.28	0.32	0.25
DASR	0.31	0.30	0.24	0.26	0.22
DWT-Dif	0.25	0.25	0.21	0.23	0.19
DWT-SWT	0.28	0.27	0.22	0.24	0.20
Proposed	**0.38**	0.36	**0.32**	**0.37**	**0.28**

The above four criteria can be used to compare the performances of different techniques for a single image. However, they are not straightforward for collectively evaluating the performance for a number of testing images. This chapter proposes four normalised criteria to better measure the improved performance of a considered resolution enhancement technique compared with a reference technique, which includes ratio of PSNR (RPSNR), ratio of RMSE (RRMSE), ratio of Entropy (RENTROPY), ratio of SSIM (RSSIM). They can be calculated as

$$RPSNR(m_1, m_2) = \frac{PSNR(m_1) - PSNR(m_2)}{PSNR(m_2)} \times 100\% \tag{5-9}$$

$$RRMSE(m_1, m_2) = \frac{RMSE(m_1) - RMSR(m_2)}{RMSE(m_2)} \times 100\% \tag{5-10}$$

$$RENTROPY(m_1, m_2) = \frac{E(m_1) - E(m_2)}{E(m_2)} \times 100\% \tag{5-11}$$

$$RSSIM(m_1, m_2) = \frac{SSIM(m_1) - SSIM(m_2)}{PSNR(m_2)} \times 100\% \tag{5-12}$$

where m_1 is the considered resolution enhancement technique and m_2 is the reference technique, which was chosen as the *bicubic* interpolation method in this chapter due to its popularity. The higher the RPSNR and RSSIM, the better the performance of the considered technique is. A positive value of RPSNR and RSSIM indicates a better performance than the reference method. The lower the RRMSE and RENTROPY, the better the performance of the considered technique is. A negative value of RRMSE and RENTROPY indicates a better performance than the reference method.

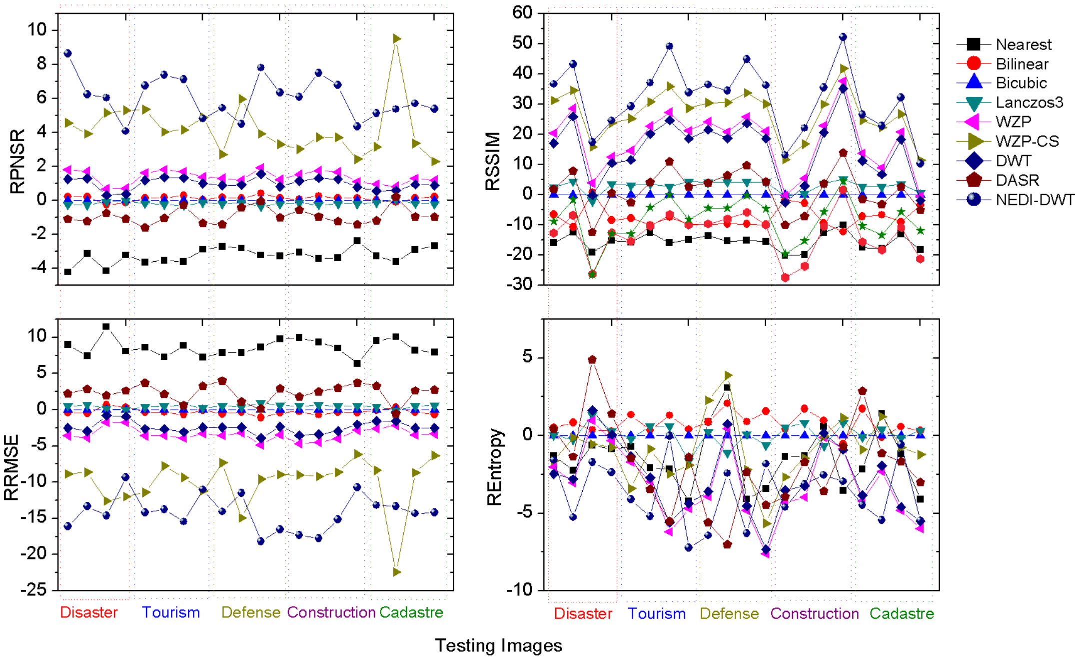


Figure 5-10 Comparison of RPSNR, RRMSE, RSSIM, and RENTROPY results of all tested images for resolution enhancement from 128×128 to 512×512 by the proposed approach and the conventional interpolation and state-of-the-art resolution enhancement techniques.

Figure 5-10 shows the results of the new proposed criteria for all 20 testing images, organised by five groups, using the selected nine resolution enhancement techniques, which include the nearest neighbour, the bilinear interpolation, the bicubic interpolation, the Lanczos interpolation, WZP, WZP-CS, DWT, DASR and the proposed technique. The wavelet function db.9/7 was used in the wavelet-based techniques. To evaluate the overall performance of the proposed technique for different classes of satellite images, Table 5-9 shows the percentage of images in which the DWT-NEDI technique has the best performance among the considered techniques. It can be inferred from Figure 5-10 that the DWT-NEDI technique has superior performance against other methods in terms of RPSNR and RRMSE, which is supported by it tops the performance for 80% of images, especially for the images from the *construction* group in which it tops 100%. In terms of RSSIM, it tops the performance for 100% of all images from five groups. For the remaining 20% of images, the WZP-CS has the best performance. However, in terms of

RENTROPY, the difference of performance is not significant among the considered methods. On average, the DWT-NEDI technique still has the best performance as it tops 65% of images.

Table 5-9 Percentage of tested images where the proposed technique tops the performance.

Image Group	RPSNR	RRMSE	RE	RSSIM
Natural disaster	75%	75%	75%	100%
Tourism	75%	75%	75%	100%
Defences and	75%	75%	50%	100%
Construction	100%	100%	50%	100%
Cadastre and land	75%	75%	75%	100%
Average	80%	80%	65%	100%

5.4.4 Variation of Wavelet Functions

All above results from the wavelet-based resolution enhancement techniques were produced by the most widely used wavelet function db.9/7. This section discusses the prospect of the proposed approach using other wavelet functions. Previous research in Chapter 4 shows that the selection of wavelet function can affect the performance. A total of 50 wavelet functions, including db.1-20, sym2-20, bior1-6 and coif1-5 [133], were tested using the proposed technique. Table 5-10 shows the calculated PSNR results for randomly selected satellite images from each group using nine wavelet functions, which include db1, db2, sym16, sym20, ciof1, ciof2, db.9/7, bior5.5, and bior6.8 as well as bicubic interpolation. These functions were selected due to their better performance than the remaining functions. The patterns for five groups of images are very similar, for example, a) the performance of the proposed technique using the selected wavelet functions is better than the *bicubic* interpolation; b) the function db.9/7 has relatively high PSNR and is in the top 3; and c) the function bior5.5 has the highest PSNR value, although its superiority over db.9/7 is relatively small.

Table 5-10 PSNR values produced by the proposed approach using the selected wavelet functions for randomly selected images from each group, where top five are highlighted.

Wavelet functions	Natural disaster	Tourism	Defences	Construction	Cadastre and land
bicubic	28.49	24.66	24.13	20.09	17.33
db1	29.99	25.67	25.16	21.26	18.45
db2	30.43	25.99	25.43	21.56	18.76
sym16	**30.60**	**26.15**	25.56	**21.74**	**18.90**
sym20	**30.60**	**26.15**	25.56	**21.74**	**18.90**
ciof1	30.57	26.11	25.52	21.69	18.86
ciof2	29.47	25.26	24.80	20.78	18.06
db.9/7	**30.70**	**26.24**	**25.63**	**21.82**	**19.00**
bior5.5	**30.74**	**26.30**	**25.69**	**21.88**	**19.05**
bior6.8	**30.61**	**26.14**	**25.56**	**21.73**	**18.90**

5.5 Summary

A new resolution enhancement approach based on DWT and NEDI was proposed in this chapter to correct the errors in image geometry and recover the details of directional high-frequency sub-bands. The observed image is decomposed into four frequency sub-bands through DWT, and then the three high-frequency sub-bands and the observed image are processed with NEDI. *To* A nonlinear adaptive thresholding process is also included to boost the edges and reduce the noise in the estimated high-frequency sub-bands for enhancing satellite images. Finally, the enhanced image is reconstructed by applying inverse DWT. The motivation for this approach is to better preserve the edges and remove potential noise in the estimated high frequency sub-bands since a direct interpolation through interpolation methods will blur the areas around edges. Five groups of satellite images (totally 20 images), randomly selected from a public data set, were tested by the proposed approach, and the results were compared with the conventional interpolation methods and state-of-the-art wavelet-based techniques. Four new criteria were introduced, aiming to better evaluate the overall performance of the proposed technique for multiple images. Results show that the proposed method outperforms conventional interpolation methods, in both objective and subjective terms, and in most scenarios, it also outperforms the state-of-the-art methods operating in the wavelet-domain.

6 Global Motion based Video Super-Resolution Reconstruction using Discrete Wavelet Transform

6.1 Introduction

High resolution (HR) images and videos are highly desirable, and strongly demanded for most digital imaging applications, not only for offering better visualisation, but also for extracting additional details. However, HR images are not always available since the setup of high resolution imaging can be expensive and the inherent physical limitations of the sensors, optics manufacturing technology, data storage and communication bandwidth. Therefore, it is essential to find an effective way in image processing to increase the resolution level at a low-cost manner, without replacing the existing imaging system. To address this challenge, the concept of super-resolution (SR) has now been sought after. This technique aims to produce a single HR image, or HR video, from a set of different successive low-resolution (LR) images captured from the same scene in order to overcome the limitations and/or possibly ill-posed conditions of the imaging system [1]. Due to its wide applications, it has been an active research over the last two decades in various fields, such as satellite imaging, medical imaging, forensic imaging and video surveillance systems.

Most SR methods consist of two main parts: image registration and image reconstruction. Image registration aims to estimate the motion between the LR images, while image reconstruction aims at combining the registered images in order to reconstruct the HR image [15]. In image registration, the motion between the reference LR image and its neighbouring LR images is required to be estimated accurately to reconstruct correctly a super-resolved image [116], [14]. When the camera is moving and the scene is stationary, global motion occurs. On the contrary, when the camera is fixed and the scene is moving, non-global (local) motion occurs. This chapter primarily focuses on the first scenario.

Registration methods can be conducted either in the spatial-domain or the frequency-domain of images. Frequency-domain methods are usually limited to global motion models, whereas spatial-domain methods generally allow more general motion models. Global models describe the motion over the entire visual field, while local models limit the motion in the neighbourhood of a pixel [93]. In the frequency-domain, Vandewalle et al. [14] presented an image registration algorithm to accurately register a series of LR images, based on the advantage of discarding their high-frequency components, where aliasing may have occurred. A planar motion model was used to estimate the shift and rotation between the images when a set of images is acquired within a short amount of time using a small camera motion. Vandewalle's method performs better than the other frequency-domain registration methods [14], such as Marcel et al. [90] and Luchese and Cortelazzo [91]. In the spatial-domain, Keren et al. [94] developed an iterative planar motion estimation algorithm that uses different filtered, down-sampled versions of the images to estimate the shift and rotation parameters based on Taylor series expansions. The goal of this pyramidal scheme is to increase the accuracy for estimating large motion parameters. Keren's method is the most accurate for sub-pixel image registration in comparison to other spatial-domain algorithms. Keren's method and Vandewalle's method have been well accepted to tackle global motions. However, the existing sub-pixel image registration methods become inaccurate when the motion is non-global.

Image reconstruction methods can also be classified into frequency-domain-based and spatial-domain-based approaches. Frequency-domain-based methods, such as approaches proposed by Tsai and Huang [20], Kim et al. [23], are usually theoretically simple and have high computationally efficiency but they are limited to only global translational motion and linear space-invariant blur during image acquisition process. Spatial-domain-based methods include, Nonuniform interpolation methods [25]- [30], Iterative back-projection (IBP) methods [31]- [35], Projection onto convex sets (POCS) methods [36]- [41], Regularized-based methods which include Maximum likelihood (ML) methods [42], [43] and Maximum a posteriori (MAP) methods [44]- [49], and an extension of this approach, called Hybrid (ML/MAP-POCS) method [50]. This kind of methods can accommodate both global and non-global motions, linear space-variant blur and noise during imaging process.

In addition to the frequency and spatial-based domains, efforts have been made using the wavelet-domain. The wavelet transform (WT) is a powerful tool that divides an image data into low and high frequency sub-bands, each of which is studied independently with a resolution corresponded to its scale [17]. The mechanism behind the strategy of WT is that the features of the image at different scales can be separated, analysed and manipulated such that global features can be examined at coarse scales, while local features can be analysed at fine scales [18]. Manipulating wavelet coefficients in sub-bands containing high-frequency components is the essential target of wavelet-based methods to solve the SR reconstruction problem. The existing literature on WT-based methods is in both the single frame (image) case and multi-frame (video) case. For the multi-frame, Izadpanahi and Ozcinar [75] presented a SR technique using DWT and bicubic interpolation of the LR video frames. Anbarjafari et al. [76] proposed a SR technique for the LR video sequences using DWT and stationary wavelet transform (SWT). However, these available methods have limited performance for a variety of noise levels, motion levels, wavelet functions, and the number of used frames.

Wavelet transforms have been tremendously successful in image de-noising applications. Xiang et al. [144] proposed wavelet-based image regularization framework by connection of concepts from structured dyadic-tee complexity measures, wavelet shrinkage, morphological wavelets, and smoothness regularization, in which edge preservation in image de-noising is controlled by a single parameter.

Different from the existing super-resolution reconstruction approaches working under either the frequency-domain or the spatial-domain, this chapter proposes a robust video super-resolution technique, based on both frequency and spatial-domains by combining the so-called DWT-NEDI and a soft-thresholding for increasing the spatial resolution and recovering the noiseless high-frequency components of the observed noisy low-resolution video sequences containing global motion, which integrates merits from methods of image registration and reconstruction in both the frequency-domain and spatial-domain. An iterative planar motion estimation algorithm between a reference frame and its neighbouring frames followed by a structure-adaptive normalised convolution (SANC) reconstruction method are applied to produce the estimated low-frequency sub-band. The discrete wavelet transform (DWT) process is employed to decompose the input reference LR frame into four sub-bands, and then the new edge-directed interpolation (NEDI) method is used to interpolate each of the high-frequency sub-bands. To better preserve the edges and remove potential noise, a nonlinear soft thresholding process is introduced to the estimated high-frequency sub-bands. Finally, the super-resolved frame is reconstructed by combining both estimated low and high-frequency sub-bands through the invert DWT.

The application of the proposed SR technique is particularly useful when the camera is moving and the observed scene is stationary. One of the motivations of the proposed technique is to provide flexibility for a variety of motion levels, noise levels, wavelet functions, and sufficient number of used LR frames since the existing wavelet-based SR methods have limited performance capabilities for these various factors and this potential has not yet been fully explored. The performance of this technique is tested on three well-known videos. The robustness of the proposed algorithm is then evaluated through empirical tests with various motion levels, noise levels, wavelet functions, and the number of used frames. Most of the existing wavelet-based SR methods have limited discussion for these various factors.

6.2 State-of-The-Art Super-Resolution Methods

6.2.1 Keren Method

Keren et al. [94] developed the following registration method, which has been found to be the most precise for registering a sequence of observed LR images captured from a moving camera with sub-pixel accuracy in respect to translation and rotation.

The relation between the horizontal shift a, the vertical shift b, and the rotation angle θ between the LR images f and g can be written as

$$g(x,y) = f(x\cos(\theta) - y\sin(\theta) + a, y\cos(\theta) + x\sin(\theta) + b) \qquad \textbf{(6-1)}$$

Expanding $\sin(\theta)$ and $\cos(\theta)$ to the first two terms in their Taylors series expansion gives the following equation

$$g(x,y) \approx f(x + a - y\theta - x\theta^2/2\,, y + b + x\theta - y\theta^2/2) \qquad \textbf{(6-2)}$$

Expanding f to the first term in its Taylors series gives the first-order equation

$$g(x,y) \approx f(x,y) + (a - y\theta - x\,\theta^2/2)\frac{\partial f}{\partial x} + (b + x\theta - y\,\theta^2/2)\frac{\partial f}{\partial y} \qquad \textbf{(6-3)}$$

The error between g and f after translation by a and b and rotation by θ can thus be approximated by

$$E(a,b,\theta) = \sum \left[f(x,y) + (a - y\theta - x\,\theta^2/2)\frac{\partial f}{\partial x} + (b + x\theta - y\,\theta^2/2)\frac{\partial f}{\partial y} - g(x,y) \right]^2 \qquad \textbf{(6-4)}$$

where the summation is over the overlapping area of f and g.

The minimum of $E(a,b,\theta)$ can be obtained by computing its derivatives with respect to a, b and θ and setting them to zero. The difference between g and f warped by (a,b,θ) will be minimised by solving the following equation for $a, b,$ and θ:

$$\begin{gathered}
\left[\sum \left(\frac{\partial f}{\partial x}\right)^2\right] a + \left[\sum \frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right] b + \left[\sum R\frac{\partial f}{\partial x}\right]\theta = \sum \frac{\partial f}{\partial x}(f - g) \\
\left[\sum \frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right] a + \left[\sum \left(\frac{\partial f}{\partial y}\right)^2\right] b + \left[\sum R\frac{\partial f}{\partial y}\right]\theta = \sum \frac{\partial f}{\partial y}(f - g) \qquad \textbf{(6-5)} \\
\left[\sum R\frac{\partial f}{\partial x}\right] a + \left[\sum R\frac{\partial f}{\partial y}\right] b + \left[\sum R^2\right]\theta = \sum R(f - g)
\end{gathered}$$

where $= x\frac{\partial f}{\partial y} - y\frac{\partial f}{\partial x}$.

By solving this set of linear equations, the motion parameters $a, b,$ and θ can then be computed. These equations were obtained using approximations which are valid only for small values of (a, b, θ).

6.2.2 Vandewalle Method

Vandewalle et al. [14] proposed a frequency-domain registration algorithm to precisely align a sequence of aliased images, based on their low-frequencies, thereby aliasing its free part. They employed a planar motion parallel to the image plane to estimate the shift and rotation parameters between the reference LR image and each of the neighbouring LR images. This motion can be represented as a function of three parameters: horizontal and vertical shifts, Δx_1 and $\Delta x 2$, and a planar rotation angle $\emptyset$. Assume that a reference image is $f_1(x)$ and its shifted and rotated version is $f_2(x)$, the relation between the images can be expressed as

$$f_2(x) = f_1\big(R(x + \Delta x)\big) \tag{6-6}$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\Delta x = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix}$, $R = \begin{bmatrix} \cos\emptyset & -sin\emptyset \\ \sin\emptyset & \cos\emptyset \end{bmatrix}$

The Fourier transform of $f_2(x)$ can be expressed as

$$\begin{aligned} F_2(u) &= \iint f_2(x) e^{-j2\pi u^T x}\, dx \\ &= \iint f_1\big(R(x + \Delta x)\big) e^{-j2\pi u^T x}\, dx \qquad \text{(6-7)} \\ &= e^{j2\pi u^T \Delta x} \iint f_1\big(R(\acute{x})\big) e^{-j2\pi u^T \acute{x}}\, d\acute{x} \end{aligned}$$

where $F_2(u)$ is the Fourier transform of $f_2(x)$ and $\acute{x} = x + \Delta x$ is the coordinate transformation. After another transformation $\acute{\acute{x}} = R\acute{x}$, the relation between the amplitudes of the Fourier transforms $|F_1(u)|$ and $|F_2(u)|$ can be computed as

$$\begin{aligned}|F_2(u)| &= \left|e^{j2\pi u^T\Delta x}\iint f_1(R\acute{x})e^{-j2\pi u^T\acute{x}}d\acute{x}\right| \\ &= \left|\iint f_1(R\acute{x})e^{-j2\pi u^T\acute{x}}d\acute{x}\right| \\ &= |F_1(u)|\end{aligned} \tag{6-8}$$

where $|F_2(u)|$ is a rotated version of $|F_1(u)|$ using the same rotation angle $\emptyset$ used in the spatial-domain rotation . As the shift values Δx only affect the phase values of the Fourier transforms $F_1(u)$ and $F_2(u)$, the amplitudes of the Fourier transforms $|F_1(u)|$ and $|F_2(u)|$ do not rely on the spatial-domain shifts. Therefore, the rotation angle can first be estimated from $|F_1(u)|$ and $|F_2(u)|$ and then the shifts can be estimated from the phase difference between $F_1(u)$ and $F_2(u)$.

6.2.3 Structure-Adaptive Normalized Convolution (SANC) Method

Pham et al. [29] proposed a so-called structure-adaptive normalised convolution (SANC) reconstruction method for fusion of irregularly sampled data obtained from the different LR images. The method is based on the normalised convolution (NC) method [30], which approximates the local signal from projections onto a set of polynomials basis functions. Using the polynomials basis functions in the NC makes it equal to a local Taylor series expansion. The intensity value at position $s = \{x + x_0, y + y_0\}$ is estimated using a polynomial expansion at the center of analysis $s_0 = \{x_0, y_0\}$

$$\hat{f}(s, s_0) = P_0(s_0) + P_1(s_0)x + P_2(s_0)y + P_3(s_0)x^2 + P_4(s_0)xy + P_5(s_0)y^2 + \cdots \tag{6-9}$$

where $\{x, y\}$ are the local coordinates samples with respect to s_0 . $P(s_0) = [P_0 P_1 \ldots P_m]^T(s_0)$ are the projection coefficients onto their entire polynomial basis functions at s_0.

An applicability function is used in NC to localise a polynomial approximation, which gives different weights to all data points in a neighbourhood in the image.

In addition, the signal certainty is introduced in NC to ensure each input signal owing its certainty value and reduce the impact of outliers caused by missing samples or erroneous registration.

6.3 Proposed Video Super-Resolution Approach

Recovering the missing high-frequency details of the given LR frames is the fundamental target of the video SR methods. The first step is sub-pixel image registration that aims to estimate the motion parameters between the reference LR frame and each of the neighbouring LR frames. When the camera is moving and the scene is stationary, global motion occurs including translation and rotation. In this work, Keren's method [94] is selected for global motion estimation which is one of the most accurate methods for sub-pixel image registration in the spatial-domain.

For image reconstruction, conventional interpolation methods (e.g., nearest neighbour, bilinear and bicubic) address the problem of reconstructing a HR image from the available LR image. However, these methods generally yield images with blurred edges and undesirable artefacts because they do not use any information pertinent to the edges in the original image. Therefore, the wavelet-based method is applied to preserve the high-frequency details (i.e. edges) and consequently construct the HR image from the given LR image. In the proposed technique, the discrete wavelet transform DWT is employed to isolate and preserve the high-frequency components of the image, and then the interpolation is applied to the high-frequency sub-bands. This is because the interpolation of isolated high-frequency components in the high-frequency sub-bands preserving more edges of the image than using a direct interpolation. A number of DWT-based interpolation methods [69]-[72] have been developed to preserve the high-frequency components in the interpolated sub-bands. Nevertheless, the blurring effect from the employed interpolation method causes the potential loss of edges in these sub-bands. Dual-tree complex wavelet transform (DT-CWT)-based interpolation methods [68], [74] and [77] have been applied to address this problem by utilising an alternative interpolation method. Jagadeesh and Pragatheeswaran [68] used edge-directed interpolation (EDI) [86] as an alternative interpolation method of high-frequency sub-bands produced by DT-CWT. Later, this method was extended by Izadpanahi and Demirel [74] for video SR. Recently, the same authors applied new-edge directional interpolation (NEDI) [87] method to better preserve the edges of the interpolated high-frequency sub-bands generated by DT-CWT for local motion-based video SR [77]. However, none of these existing wavelet-based methods have tackled the problem of noisy high-frequency details corrupted by the limitations of imaging systems. For the current work, a combination of DWT, NEDI and a soft thresholding process is proposed not only for preserving the high-frequency details, but also for recovering the noiseless high-frequency information. One-level DWT process decomposes the input LR reference frame into four frequency sub-bands (LL, LH, HL and HH) in the frequency-domain. The high-frequency sub-bands (LH, HL and HH) are interpolated using the NEDI method with the scale factor α. Generally, real video sequences are most commonly contaminated by noise, such as additive Gaussian noise. Therefore, to better preserve the edges and remove potential noise in the estimated high-frequency sub-bands, a nonlinear thresholding procedure that uses a soft-thresholding technique [141] is applied to process the estimated wavelet coefficients. The universal threshold τ for the considered sub-band can be calculated as

$$\tau = \sigma\sqrt{2\log(N)/N} \tag{6-10}$$

where σ is the standard deviation of the sub-band and N is the total number of pixels. The nonlinear soft-thresholding function is defined as

$$X_{out}(i,j) = \begin{cases} X_{in}(i,j) - \tau & X_{in}(i,j) > \tau \\ 0 & |X_{in}(i,j)| \leq \tau \\ X_{in}(i,j) + \tau & X_{in}(i,j) < -\tau \end{cases} \tag{6-11}$$

Eq. (6-10) is chosen in the proposed method considering the prospect of automation in the proposed method and successful application of this equation in similar studies [141].

The rationale to include this thresholding process is that the signal energy is often concentrated on a few coefficients while the noise energy is spread in all coefficients in the wavelet domain. Therefore, the nonlinear soft-thresholding tends to maintain few larger coefficients representing the signal and reduce noise coefficients to zero in the wavelet domain. On the other hand, in the spatial-domain, when the LR frames are precisely registered by Keren’s method, the registered frames can be combined to reconstruct the missing high-resolution information and produce the low-frequency sub-band. In this work, structure- adaptive normalised convolution (SANC) reconstruction method [29] is applied, with half of the scale factor α/2. This algorithm is used for fusion of irregularly sampled LR frames to recover the high-frequency details and generate the estimated LL sub-band, as the LL sub-band produced by the DWT does not contain any high-frequency information. Finally, inverse DWT (IDWT) process is applied to achieve a super-resolved frame by combining the estimated LL sub-band and processed high-frequency sub-bands. The block diagram of the proposed video SR technique is illustrated by Figure 6-1.

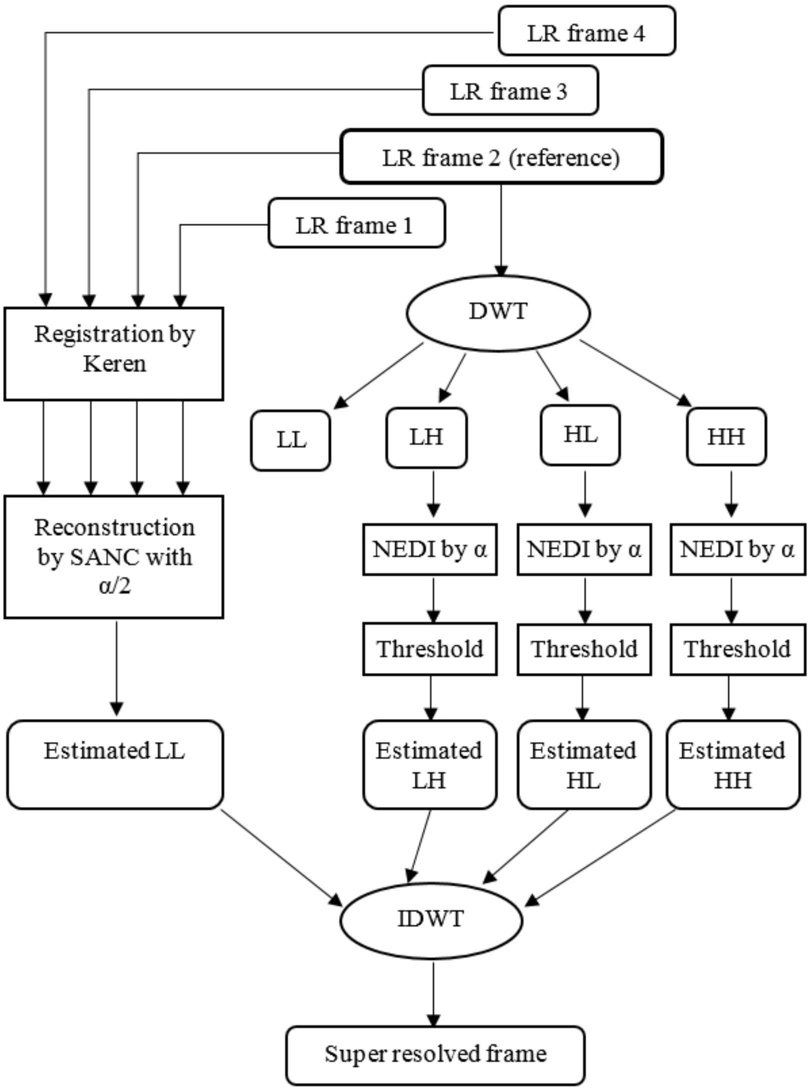


Figure 6-1 Block diagram of the proposed video SR technique

The combination of DWT with NEDI aims to recover the edge details of directional high-frequency sub-bands and decrease the undesirable inter-directional interference in the SR process. This merit cannot be achieved using only the NEDI method, as indicated in the results section. The application of this soft-thresholding function is based on the hypothesis that the large coefficients in the high-frequency sub-bands reflect the true edges of objects, while the small coefficients reflect the noise, which is demonstrated by Figure 6-2. Figure 6-2(a) shows the reconstruction image of high-frequency sub-bands only using IDWT without thresholding. Both the true edges and noise can be clearly observed. Figure 6-2(b) shows the reconstruction image of high-frequency sub-bands only with thresholding where the small coefficients are removed. It can be observed that the noise is significantly reduced, particularly in the background, while most of the true edges of the human body are preserved. The reconstruction image of high-frequency sub-bands only where the large coefficients are removed is illustrated by Figure 6-2(c), which is dominated by noise and very few true edge information can be observed. To demonstrate the importance of this process, a region of the produced HR image using the proposed method without the soft-thresholding is shown in Figure 6-2(d), where the noise can be clearly observed.

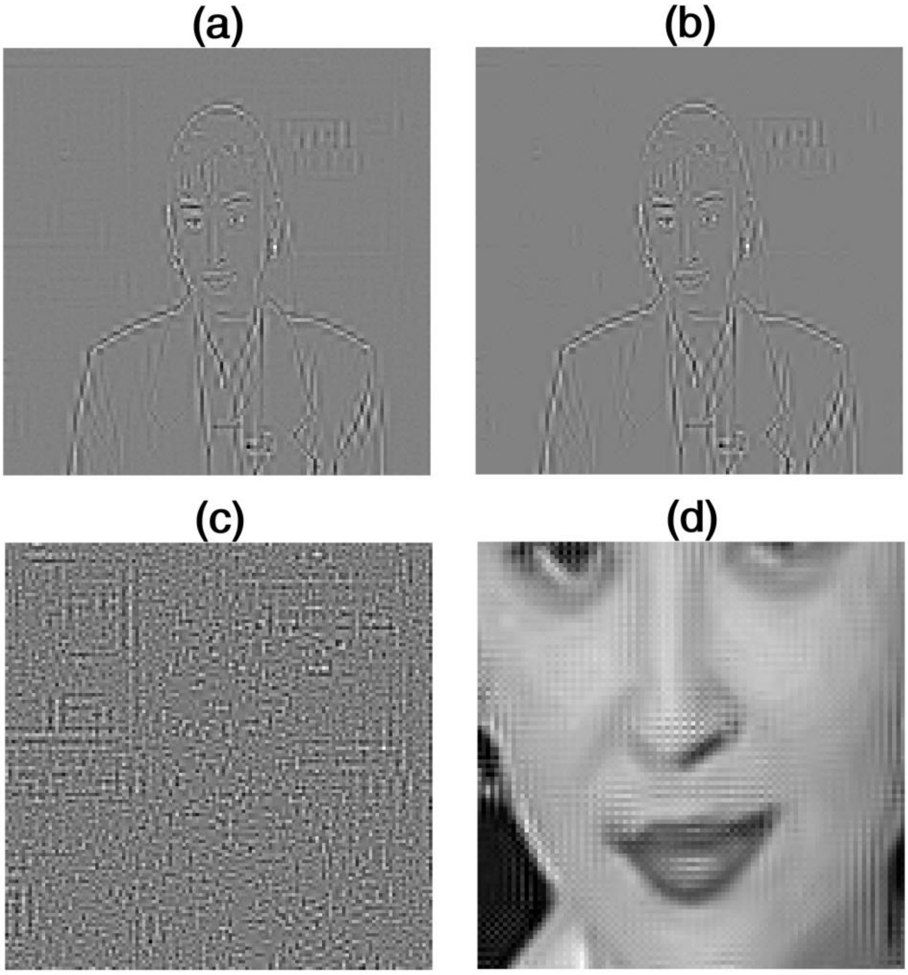


Figure 6-2 An example to help justify the use of thresholding process. (a) the reconstruction image of high-frequency sub-bands only without thresholding; (b) the reconstruction image of high-frequency sub-bands only with thresholding where the small coefficients are removed; (c) the reconstruction image of high-

frequency sub-bands only where the large coefficients are removed, (d) the produced HR image without thresholding.

The proposed technique can be summarised by the following steps:

1. Consider four consecutive frames from the LR video;
2. Estimate the motion parameters between the reference frame and each of the other LR frames using global motion estimation algorithm proposed by Keren;
3. Apply one-level DWT to decompose the input LR reference frame into four frequency sub-bands;
4. Apply the NEDI method to the LH, HL and HH high-frequency sub-bands with the scale factor of α;
5. Calculate the threshold τ for each high-frequency sub-band;
6. Apply the nonlinear soft-thresholding process for each high-frequency sub-band to create the estimated $\widehat{\mathrm{LH}}$, $\widehat{\mathrm{HL}}$ and $\widehat{\mathrm{HH}}$;
7. In the spatial-domain, employ the structure adaptive normalised convolution SANC with half of the scale factor α/2 to create the estimated $\widehat{\mathrm{LL}}$;
8. Apply IDWT using $\left(\widehat{\mathrm{LL}}, \widehat{\mathrm{LH}}, \widehat{\mathrm{HL}}, \widehat{\mathrm{HH}}\right)$ to produce the output super-resolved frame.

6.4 Results and Discussion

The proposed super-resolution technique was tested on three well-known video sequences, namely, "*Mother & daughter*", *"Akiyo*", and *"Foreman*". The video sequences were downloaded from a public data-base Xiph.org. The proposed method and other methods for comparison were implemented using Matlab 2015. The original high-resolution test videos were resized to 512×512 pixels which are considered as the ground truth to evaluate the performance of the proposed algorithm. The reason for resizing the original test videos to 512×512 pixels is for the convenience to compare the performance of the proposed SR technique with the other state-of-the-art resolution enhancement and SR reconstruction techniques in the literature. The frame rate of the test videos is 30 frames per second and each of the video sequences has 300 frames. Based on the observation model, the input LR video frames with the size of 128×128 pixels were created as follows. Each original HR video frame is (1) blurred by a low-pass filter, (2) down-sampled in both the vertical and the horizontal directions by a scale factor of 1/4, and (3) added by a Gaussian noise with a certain value of signal-to-noise ratio (SNR).

6.4.1 Visual and Quantitative Performance Evaluation

This example aims to evaluate the overall performance of the proposed technique with a typical selection of parameters against other methods. For each original HR frame, four shifted and rotated LR frames were generated and down-sampled, and a Gaussian noise was then added with the SNR value of $30\ \mathrm{dB}$. The motion vectors were randomly produced with a standard deviation (STD) of 2 for shift of both directions and 1 for rotation. The wavelet function was chosen as db.9/7.

Figure 6-3 Results of produced HR images using different SR methods for a randomly selected frame of *Akiyo, Mother & daughter,* and *Foreman* video sequence respectively. Column 1: input LR frame, Column 2: Bicubic, Column 3:

Vandewalle-SANC, Column 4: Keren-SANC, and Column 5: the proposed technique.

Figure 6-3 shows the super-resolved frames using the proposed method and other methods, for selected arbitrarily regions from the video sequences *Akiyo*, *Mother & daughter*, and *Foreman*, respectively. It can be observed that the proposed technique produces the best visual quality in terms of preserving the edges, and removing the noise and aliasing artefacts in comparison to the other considered methods. The proposed technique preserves more information of the edges of the original HR video frame without smearing. For example, the edges of the face in *Akiyo* produced by the proposed method are much cleaner in comparison with the images produced using other methods. Similar visual results can be observed for other edges in *Akiyo* and for other tested videos. Additionally, the noise and aliasing artefacts are removed by the proposed method in comparison to the other methods. For example, the aliasing artefacts in *Mother* shoulders and hands are removed substantially by the proposed technique as well as the Gaussian noise on the face of *Foreman*, while these aliasing artefacts and noise are clearly presented in the images produced by the other methods. From the motion estimation point of view, the aliasing high-frequency components due to down-sampling process appear to have different motions than the low-frequency components, and cause incorrect motion estimation [106]. Moreover, a larger noise level generates errors in motion estimation.

To further investigate the improvement of the proposed method, Figure 6-4 shows the local PSNR maps for different scenarios of the example *Akiyo*. The local PSNR map was calculated by a 5×5 pixels window. Figure 6-4(a) shows the PSNR distribution between the raw HR image and the interpolated HR image produced by nearest neighbour method, which indicates the location of noise introduced by the degrading process. Regions of the human body have more noise (blue regions) introduced by the degrading process, while background regions have less noise (yellow regions). Blue regions tend to be areas with fine features (like boundary of the human body), while yellow regions tend to have more coarse structures. Similar representation using the proposed method is shown in Figure 6-4(b), inspection of which proves that both background and fine features have been better recovered. To break down the contribution of each component, Figure 6-4(c) and (d) show the PSNR gain of the proposed method over NEDI and Keren-SANC, respectively. It can be observed that fine features are significantly improved in comparison to NEDI due to the consideration of adjacent frames, while the improvement of coarse structures is relatively small. Coarse structures are significantly improved in comparison to Keren-SANC, while the improvement of fine features is relatively small. All these observations clearly demonstrate that the proposed method improves the quality of both background and true edges, but other methods can only have one merit.

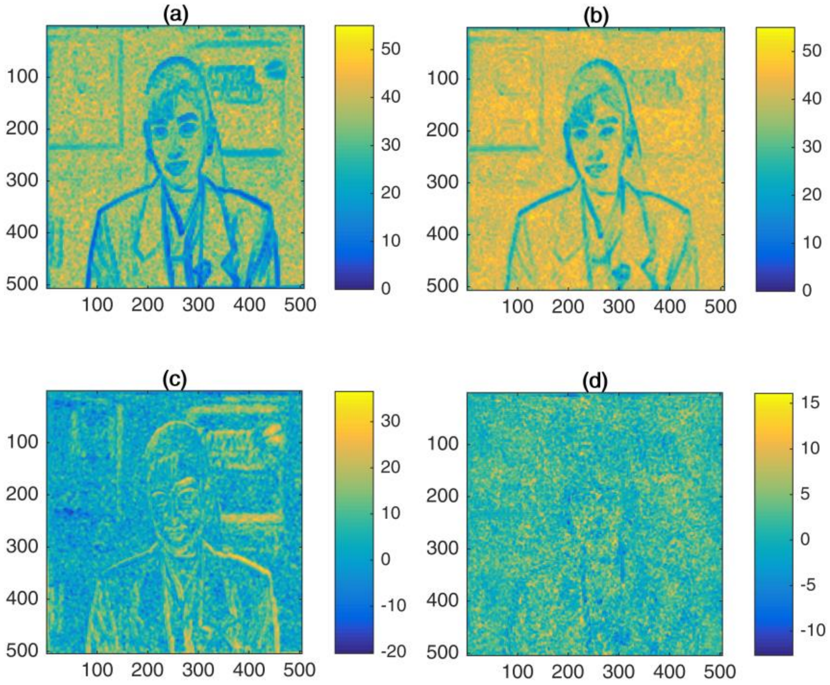


Figure 6-4 An example to show the improved distribution of the proposed technique. (a) The distribution of PSNR between the HR image using the nearest interpolation method and the raw image; (b) the distribution of PSNR between the super-resolved image using the proposed method and the raw image; (c) the PSNR gain between the proposed method and NEDI; (d) the PSNR gain between the proposed method and Keren-SANC

For quantitative evaluation of the experimental results, the nearest neighbour and bicubic interpolation methods, state-of-the-art resolution enhancement methods including NEDI [87], DASR [70], DWT-Diff [71], DWT-SWT [72], and state-of-the-art SR methods Keren-SANC and Vandewalle-SANC were implemented to compare the performance of the proposed technique.

Table 6-1 shows the comparison of the averaged PSNR and SSIM results of 100 frames produced from the proposed method and the other considered methods on the test videos. The reason to choose the first 100 frames of tested video sequences is that there is no big difference in results in terms of the content of video frames based on global motion. It is clearly shown that the proposed method achieves the highest average PSNR and SSIM values ($31.48\ \mathrm{dB}, 30.57\ \mathrm{dB}$ and $23.88\ \mathrm{dB}$ for PSNR respectively; $0.90, 0.91$ and 0.84 for SSIM) for three tested videos. Achievement of this improved performance is probably because that the DWT-based SR reconstruction approach is more effective to recover the noiseless high-frequency components of the given LR frames, where the true edges are preserved and the noise is removed benefiting from the nonlinear soft thresholding method. Additionally, the combination of DWT and NEDI enables the recovering of the edge details of directional high-frequency sub-bands and reduces the annoying inter-directional interference in the SR process. For the videos *Mother & daughter* and *Akiyo*, the proposed technique based on global motion of the entire frame is suitable for these videos and produces better PSNR results than the interpolation, resolution enhancement and classic SR methods (16% and 11% increment over Keren-SANC respectively). For the video *Foreman*, although the PSNR result achieved by the proposed method is higher than the other considered methods (17% increment over KEREN-SANC), the performance gain can be further increased by utilising local motion which constrains the motion in the neighbourhood of a pixel by dividing each frame into multiple blocks and processing each block individually. Similar observation can be concluded based on SSIM values.

Table 6-1 The averaged PSNR and SSIM values of 100 frames produced from different methods for three tested videos

SR methods	Mother& Daughter		Akiyo		Foreman	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Nearest	24.80	0.69	24.54	0.75	20.89	0.61
Bicubic	25.92	0.77	25.77	0.82	22.02	0.72
NEDI [94]	24.88	0.75	24.89	0.81	21.19	0.71
DASR [70]	23.76	0.62	23.83	0.69	20.16	0.55
DWT-Diff [71]	22.57	0.55	22.58	0.62	18.97	0.47
DWT-SWT [72]	23.10	0.57	23.09	0.65	19.83	0.50
VAN-SANC	22.05	0.75	25.18	0.81	19.96	0.72
Keren- SANC	27.14	0.82	27.51	0.84	20.41	0.70
Proposed method	**31.48**	**0.90**	**30.57**	**0.91**	**23.88**	**0.84**

6.4.2 Performance for Variety of Noise Levels

To demonstrate the robustness of the proposed method against noise benefiting from the adaptive thresholding process, four shifted and rotated LR frames for each original HR frame were generated and down-sampled, and the motion vectors were randomly produced with a standard deviation of 2 for shift and 1 for rotation. The wavelet function was chosen as db.9/7. The noise level was increased from $50\,\mathrm{dB}$ to $20\,\mathrm{dB}$ with the step of $5\,\mathrm{dB}$. The first 10 frames from *Akiyo* video were tested by the proposed method and other different methods, and the results were averaged. To demonstrate the robustness of the proposed SR method against noise and other parameters, only the first 10 frames were tested and the results were averaged. Results have no significant difference if more frames are considered.

Table 6-2 shows the comparison of averaged PSNR results between different methods of the first 10 frames. The last column shows the PSNR increment over Keren-SANC in percentage by the proposed technique. It can be clearly seen that the proposed method consistently has the best performance for every noise level. Furthermore, the performance is even better for images corrupted by high noise level (15%, 21%, 18% increment for $30\,\mathrm{dB}$, $25\,\mathrm{dB}$ and $20\,\mathrm{dB}$ respectively) than those with low noise level (10%, 8%, 5%, 3% increment for $50\ \mathrm{dB}$, $45\ \mathrm{dB}$, $40\ \mathrm{dB}$ and $35\ \mathrm{dB}$ respectively).

Table 6-2 The averaged PSNR results of 10 frames from Akiyo test video for each noise level, range from 20 dB to 50 dB with 5 dB step

SNR	Nearest	Bicubic	Keren-SANC	Proposed method	Increment
50 dB	25.20	26.50	28.18	31.13	10.47%
45 dB	23.86	24.85	28.46	30.82	8.29%
40 dB	24.98	26.29	30.05	31.64	5.29%
35 dB	24.18	25.20	30.49	31.30	2.65%
30 dB	23.84	24.86	26.50	30.37	14.60%
25 dB	23.69	24.85	24.80	30.08	21.29%
20 dB	23.42	24.93	24.31	28.63	17.78%

6.4.3 Performance for Variety of Wavelet Functions

The above results from the proposed technique were produced by the most widely used wavelet function db.9/7 in image SR applications. This section discusses the prospect of the proposed approach using other wavelet functions. Previous research in Chapter 4 shows that the selection of wavelet function can affect the performance of SR techniques. In this experiment, the same parameters were chosen except that the noise level was fixed as 30 dB, and the wavelet function is variable. Table 6-3 shows the averaged PSNR and SSIM results of the first 10 frames for the *Akiyo* video sequence produced by the proposed method using nine wavelet functions, which include db1, db2, sym16, sym20, ciof1, ciof2, bior4.4 (db.9/7), bior5.5, and bior6.8. The results show that the proposed technique can perform well under other wavelet functions apart from db.9/7, even better than db.9/7. Bior4.4 is equal to db.9/7 [129]. In terms of both PSNR and SSIM values, the wavelet function with top 5 performance are sym20, sym16, bior6.8, bior4.4 and coif1 respectively, although the difference between them is not significant.

Table 6-3 The averaged PSNR and SSIM values of 10 frames produced by the proposed technique for different wavelet functions

Wavelet functions	PSNR	SSIM
Db1	28.18	0.88
Db2	29.95	0.90
Sym16	**30.83**	**0.92**
Sym20	**30.89**	**0.92**
Coif1	**30.39**	**0.91**
Coif2	26.78	0.85
Bior4.4	**30.72**	**0.92**
Bior5.5	30.20	0.90
Bior6.8	**30.81**	**0.92**

6.4.4 Performance for Variety of Motion Levels

This section is dedicated to discuss the effectiveness of motion level (shift and rotation) on the performance of the proposed algorithm. In this experiment, the shift on both horizontal and vertical directions and rotation angle were randomly selected with the standard variation STD changing from 1 to 4 during generating the input LR frames from the original HR frame. Four shifted and rotated LR frames for each original HR frame were generated and down-sampled. The wavelet function was chosen as db.9/7, and the noise level was fixed as $30\ \mathrm{dB}$.

The averaged PSNR and SSIM results of the first 10 frames form *Akiyo* video produced by the proposed technique with different motion levels are shown in Table 6-4. It can be observed that the proposed method produces the highest PSNR and SSIM values when the motion level is relatively small. This is because the estimation of a small motion is usually more accurate and leads to better reconstruction of the original HR frames. When the motion level is large, the values of PSNR and SSIM drop as expected. This is because a large motion is more difficult to be measured accurately and errors in motion estimation prevent reconstructing the original HR frames correctly. From the considered smallest to largest motions, 12% and 7% decrease of PSNR and SSIM values respectively has been observed.

Table 6-4 The averaged PSNR and SSIM values of 10 frames produced by the proposed technique with different motion levels

STD of Shift	STD of Rotation					
	1		2		4	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	**30.74**	**0.92**	29.44	0.90	28.09	0.87
2	29.16	0.90	29.00	0.89	27.97	0.88
4	27.35	0.86	27.13	0.86	26.90	0.85

6.4.5 Performance for Variety of Number Frames

This section aims to evaluate the effectiveness of the number of used frames on the performance. In all previous experiments, it was assumed that the shift and rotation parameters were randomly produced. But in real applications, the camera usually moves towards one direction which means the shifts change monotonously. The shifts include two motion vectors, horizontal shift Δ_x and vertical shift Δ_y. To simplify the process, in this experiment, the rotation angle was randomly selected with the standard deviation of 1, and only the shifts are changed. The shifts were produced based on

$$\Delta_x(i) = \Delta_y(i) = \frac{i}{N} \tag{6-12}$$

where $i(i = 1,2, \dots, N)$ denotes the time index of LR frames and N denotes the total number of used frames. All other parameters are same as the previous experiment.

Table 6-5 shows the averaged PSNR and SSIM results of the first 10 frames for the *Akiyo* video sequence produced by the proposed technique using the number frames of 4, 8, 16 and 32 respectively. It can be observed that, as expected, higher PSNR and SSIM values were achieved with more number of sampled frames. However, the increment is only about 1% when the number of frames changes from 4 to 32. This observation is because that the used motion model is simple, and there is very limited extra contribution from 32 frames in comparison to 4 frames. More improvement could be achieved when the motion is more complicated and the motion is corrupted by more noise.

Table 6-5 The averaged PSNR and SSIM values of 10 frames produced by the proposed technique by sampling different number of frames

Number of sampled frames	PSNR	SSIM
4	29.81	0.91
8	30.06	0.92
16	29.98	0.92
32	**30.07**	**0.92**

6.5 Summary

A robust video super-resolution reconstruction approach based on combining discrete wavelet transform, new edge-directed interpolation and the nonlinear soft-thresholding has been proposed in this chapter for noisy LR video sequences with global motion to recover the noiseless high-frequency details and increase the spatial resolution, which integrates properties from methods of image registration and reconstruction. The application of the proposed SR technique is especially useful for a sequence of images captured from a moving camera where images differ mostly by translation and rotation. Firstly, an iterative planar motion estimation algorithm by Keren is used to estimate the motion parameters between a reference LR frame and its neighbouring LR frames in the spatial domain. The registered frames are combined by the SANC reconstruction method to output the estimated low-frequency sub-band. Secondly, the DWT is employed to decompose each input reference LR frame into four frequency sub-bands in the frequency-domain. The NEDI is employed to process each of three high-frequency sub-bands, which are then filtered using the adaptive thresholding process to preserve the true edges and reduce the noise in the estimated high-frequency sub-bands. Finally, by combining the estimated low-frequency sub-band and three high-frequency sub-bands, a super-resolved frame is recovered through the invert DWT process.

Subjective results show that this approach can better preserve the edges and remove potential noise in the estimated high-frequency sub-bands since a direct interpolation will blur the areas around edges. Three well-known videos (totally 100 frames for each) were tested, and the quantitative results show the superior performance of the proposed method. The proposed method tops the averaged PSNR values and SSIM values ($31.48\,\mathrm{dB}, 30.57\,\mathrm{dB}$ and $23.88\,\mathrm{dB}$ for PSNR respectively; $0.90, 0.91$ and 0.84 for SSIM) of 100 frames for three videos respectively, and the averaged increment over KEREN-SANC is 16%, 11%, and 17% respectively. The performance against noise has also been analysed. Analysis based on contribution of each component clearly demonstrates that the proposed method improves the quality of both background and true edges, but other methods usually can only have one merit.

One of the motivations of this chapter is to address the limited performance capabilities of most the existing wavelet-based SR methods for a variety of motion levels, noise levels, wavelet functions and adequate number of used frames, do empirical tests and analyse how these factors can affect the performance of the proposed method. The conclusions are:

- The proposed technique has produced 10%, 8%, 5% and 3% averaged increment of PSNR for image corrupted by low level noise with the SNR value of 50 dB, 45 dB, 40 dB and 35 dB respectively. It has produced 15%, 21% and 18% averaged increment of PSNR for image corrupted by high level noise with the SNR value of 30 dB, 25 dB and 20 dB respectively.

- The proposed technique can perform well using other wavelet functions apart from db.9/7, even better than db.9/7, although the difference between them is not significant.
- Different levels of motion can affect the performance. From the considered smallest to largest motions, 12% and 7% decrease of PSNR and SSIM values respectively has been observed.
- If the motion is simple, the number of sampled frames has limited improvement on the performance due to the limited extra information. If the motion is complex and corrupted by high level of noise, significant improvement is expected using more frames. This research requires further study.

A limitation of this method is that it can only be applied to video sequences with global motion between frames. However, it can be extended to local motion by dividing the video frame into multiple blocks such that each block has some uniform motion and then applying this method to each block.

7 Video Super-Resolution based on Adaptive Local Motion Decomposition and Wavelet Transform Reconstruction

7.1 Introduction

Super-resolution (SR) reconstruction technique aims to reconstruct a single high-resolution (HR) image, or a set of HR images (video), from a sequence of observed low-resolution (LR) images captured from the same scene [1]. There are two main stages in SR algorithms: image registration and image reconstruction. Image registration aims to estimate the motion parameters between the reference LR image and its neighbouring LR images at sub-pixel accuracy, while image reconstruction aims to integrate the registered images and eliminate any degradation operators to estimate the target HR image [15]. Motion estimation/registration plays a critical role for the success of SR algorithms. When the camera shifts and the observed scene is stationary, the sub-pixel displacements for all pixels in the LR images are identical, which is called global motion. When the camera is stationary and the scene moves, the sub-pixel displacements for different regions are variant, which is called non-global or local motion. However, the existing sub-pixel motion estimation methods become inaccurate and cause errors when the motions of the objects are local. For local motions, motion detection methods have to be applied to segment regions belonging to moving objects from the rest of an image [105]. There are three main categories of these methods, including background subtraction, optical flow, and temporal differences. Background subtraction is the most popular method that aims at the discrimination of moving objects from a maintained and updated background model. It is relatively easy to implement and requires less computational complexity [100]-[102]. Although the optical flow method [106], [107] shows the projected motion on the image plane with good approximation of the complex background, it often demands very high computational complexity. Temporal differencing method can be effectively adapted to environmental changes, but it often achieves incomplete detection of the shapes of moving objects [101], [102].

Image reconstruction algorithms can be classified into frequency-domain-based and spatial-domain-based approaches. Between these two domains, the wavelet transform (WT) approach has emerged as an effective new transform over the Fourier transform for analysing real-world signals because of its attractive properties, such as locality and multi-resolution [19], [16]. Discrete wavelet transform (DWT) is one of the recent wavelet transforms and is being employed as a powerful mathematical tool in many image/video processing applications to isolate and preserve the high-frequency components of the image. DWT decomposes the given image into four frequency sub-bands using the property of dilations and translations by a single wavelet function called mother wavelet [19], [129]. However, the decimated DWT is shift variant because of the decimation process exploited in the transform and, thus, suppression of wavelet coefficients introduces artefacts into the image [62], [63]. Estimating the unknown wavelet coefficients in sub-bands containing high-frequency components is the essential target of wavelet-based methods to reconstruct the HR image from the given LR image/images. SR reconstruction techniques strongly rely on accurate motion estimation for the recovery of finer details. The need for sub-pixel accuracy has restricted its application to video sequences with global simple motions only. Unfortunately, real-world video sequences contain complex local motions and motion estimation becomes a difficult task for such videos. A variety of research has been conducted to address SR reconstruction problem for practical videos and reduce local registration errors. For local motion-based video SR, Izadpanahi and Demirel divided each frame into stationary and motion blocks for better registration. They applyed NEDI method to better preserve the edges of the interpolated high-frequency sub-bands generated by dual-tree complex wavelet transform (DT-CWT) of the LR video frames. However, they could not address directly the problem of complex-motion blocks which causes inaccurate motion estimation and artefacts around motion boundaries [77].

Probabilistic motion estimation algorithms and steering kernel regression algorithms have been proposed to circumvent the problem of SR by avoiding the explicit need for precise sub-pixel motion estimation. Protter et al. [112]

developed a non-local means (NLM) de-noising method by measuring the similarity of image patches across space-time and giving relatively higher weights to more similar patches. This method was then extended by Cheng et al. [113] based on dividing each frame into simple areas and complex areas to improve the performance of the NLM algorithm. Takeda et al. [114] introduced a method based on the extension of steering kernel regression framework to 3-D signals for performing video de-noising, spatio-temporal upscaling and SR without the need for explicit sub-pixel accuracy motion estimation.

As most real senses contain complex motions with local motions and spatial variations, super-resolving real-world video sequences remains challenging. This chapter focuses on addressing the challenge of video sequences containing complex local motions by proposing a new video SR reconstruction framework that initially divides the observed LR video frame into background regions, simple-motion blocks, and complex-motion blocks. Each of block/region is then super-resolved using appropriate registration and reconstruction methods or a single wavelet-based resolution enhancement approaches to achieve an overall better performance. This method aims to improve image quality through (a) motion decomposition for more accurate motion estimation, (b) reduction of background boundary and extension of motion blocks to reduce the artefacts caused by boundary, and (c) employment of the discrete wavelet transform to preserve the true edges of objects meanwhile compress the noise.

7.2 State-of-The-Art Background Subtraction Methods

Background subtraction is a popular approach for motion detection, because it is relatively simple to implement and produces reasonable motion segmentation particularly for situation with a relatively static background. The most commonly used methods include the simple background subtraction (SBS) [103], the running average (RA) [104], and the sigma-delta estimation (SDE) [105]. In this section, the SBS and SDE methods will be described and the RA method will be described and used in the next section for the proposed framework.

7.2.1 Simple Background Subtraction (SBS) Method

SBS discriminates moving objects by taking the absolute difference between the static background frame $B(x,y)$ and the current frame $I_t(x,y)$ from the video sequence. A binary motion detection mask is calculated by the equation:

$$D(x,y) = \begin{cases} 1, & if|I_t(x,y) - B(x,y)| > \tau \\ 0, & if|I_t(x,y) - B(x,y)| \leq \tau \end{cases} \tag{7-1}$$

where τ is an experimentally selected threshold which distinguishes pixels as either the foreground or the background in a video frame. If the absolute difference between $B(x,y)$ and $I_t(x,y)$ is bigger than τ, the pixels of the detection mask are labeled with 1, which indicate moving objects; otherwise the pixels are labelled with 0, which indicate background. However, this method demonstrates poor performance in most real video sequences. The main reason that causes the SBS method fails to respond accurately is that the noise or illumination changes occur in the incoming video frame and static objects occur in the reference frame [103].

7.2.2 Sigma-Delta Estimation (SDE) Method

An adaptive background model is generated by sigma-delta estimation (SDE) method [105] which is used to calculate the temporal statistics of the pixels in the original video sequence. For the first background estimation, the sgn function is used to evaluate the intensity of the background. The sgn function is defined as follows:

$$sgn(a) = \begin{cases} 1, & ifa > 0 \\ 0, & ifa = 0 \\ -1, & ifa < 0 \end{cases} \tag{7-2}$$

where a is an input real value intensity. The background estimation is expressed as follows:

$$B_t(x,y) = B_{t-1}(x,y) + sgn\big(I_t(x,y) - B_{t-1}(x,y)\big) \tag{7-3}$$

where $B_t(x,y)$ is the current background model, $B_{t-1}(x,y)$ is the previous background model, and $I_t(x,y)$ is the current incoming video frame. The background intensity model increases or decreases by a value of one through the sgn function evaluation at each frame. Then absolute difference $\Delta_t(x,y)$ is computed as follows:

$$\Delta_t(x,y) = |I_t(x,y) - B_t(x,y)| \tag{7-4}$$

For the time-variance $V_t(x,y)$, the computation makes use of the sgn function which measures motion activity to distinguish every pixel as background or moving object.

$$V_t(x,y) = V_{t-1}(x,y) + sgn\big(N \times \Delta_t(x,y) - V_{t-1}(x,y)\big) \tag{7-5}$$

where $V_t(x,y)$ is the current time-variance, $V_{t-1}(x,y)$ is the previous time-variance, N is the predefined parameter which ranges from 1 to 4.

According to the produced current time-variance; the binary motion detection mask $D(x,y)$ is detected by the equation

$$D_t(x,y) = \begin{cases} 1, if \Delta_t(x,y) > V_t(x,y) \\ 0, if \Delta_t(x,y) \le V_t(x,y) \end{cases} \tag{7-6}$$

7.3 Proposed Video Super-Resolution Method

7.3.1 Framework

The new method, called Adaptive Local Motion Decomposition and Wavelet Transform Reconstruction (ALMD-WTR), involves three modules: motion detection module, motion decomposition module, and reconstruction module. As illustrated by Figure 7-1, the motion detection module aims to divide the input LR frames into background blocks and motion blocks in a binary form. In the motion decomposition module, the adjacent background blocks are connected and transformed into background regions. Additionally, the motion blocks are further divided into two types of blocks: simple-motion blocks and complex-motion blocks. For the simple-motion blocks and background regions, tailored registration and reconstruction methods based on considering the reference frame and the adjacent frames are used for each type of blocks/regions in the reconstruction module to produce the HR blocks/regions.

For the complex-motion blocks, a single wavelet-based resolution enhancement approach based on the reference frame only is applied for this type of blocks to produce the HR blocks, and finally a super-resolved frame is reconstructed by merging all of them. Details of each module are discussed below.

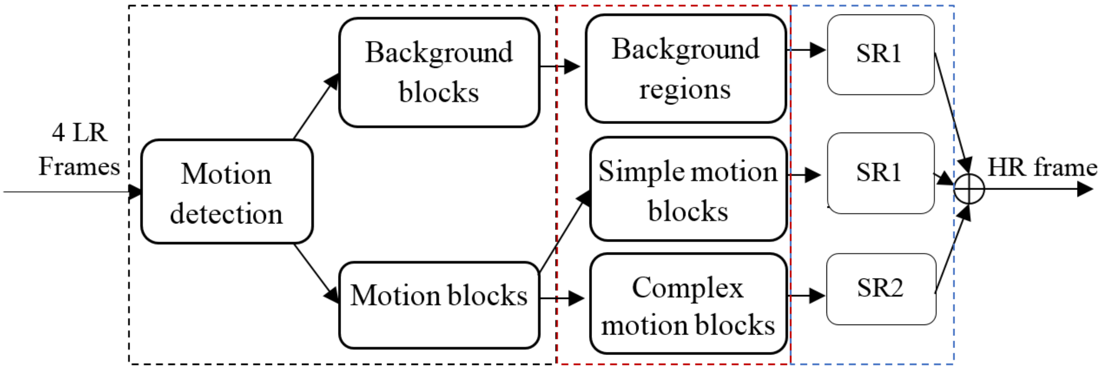


Figure 7-1 Block diagram of the proposed ALMD-WTR method, where the left rectangle shows the motion detection model, the middle one shows the motion decomposition module and the right one shows the reconstruction module.

7.3.2 Motion Detection Module

This chapter uses the background subtraction method to detect moving objects due to its advantages of easy implementation and low computational complexity. The concept of this method is building a model of the static scene without moving objects called background, and then comparing each frame in the sequence to this background to differentiate moving objects, called foreground [105]. In this section, the running average (RA) [104] is described and used in the proposed method The RA method is able to adapt for temporal changes of weather and lighting on motion detection in the video sequence by iteratively updating the background frame of the adaptive background model to guarantee reliable motion detection. Due to this prominent adaptability, this algorithm is employed in the proposed ALMD-WTR method.

In the RA method, the current background frame $B_t(x,y)$ is updated by the equation

$$B_t(x,y) = (1-\beta)B_{t-1}(x,y) + \beta I_t(x,y) \qquad \textbf{(7-7)}$$

where $B_{t-1}(x,y)$ is the previous background frame and $I_t(x,y)$ is the current considered LR frame. The updating rate β represents the speed of new changes in the scene updated to the background frame, which is usually chosen as a small value. A difference image $D_t(x,y)$ is then produced iteratively based on an absolute difference between the current frame and the current background frame

$$D_t(x,y) = |I_t(x,y) - B_t(x,y)| \tag{7-8}$$

A binary (pixel-based) motion mask $PM(x,y)$, initially each pixel of which is assigned to zero, is generated by transforming $D_t(x,y)$ into the binary form based on a threshold τ_p. A few experiments have been conducted in this study to choose the best threshold and the results showed that the Otsu's method [145] performs well in this process. The Otsu's method chooses the threshold to minimise the intraclass variance of the black and white pixels, and produce a more reasonable motion distribution. Therefore, it is used for all examples in this chapter. An example of the detected difference image and pixel-based motion mask using the video sequence of *Akiyo* is illustrated by Figure 7-2(a) and (b), where the size of observed LR images is 128×128 pixels, β has been chosen as 0.05. The foreground (white) and background (black) are clearly detected.

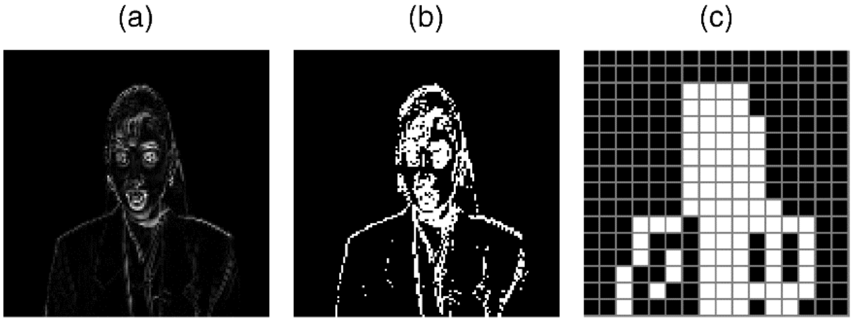


Figure 7-2 Results produced from the motion detection module, where the white colour represents the motion pixel or block and black colour represents the background pixel or block. (a) $D_t(x,y)$; (b) $PM(x,y)$; (c) $BM(x,y)$.

To construct the units for processing, the pixel-based mask $PM(x,y)$ is transformed into a block-based mask $BM(x,y)$. Firstly, the frame is divided into multiple blocks with a size of $n \times n$ pixels. Considering each block, if the block contains white pixels whose summation is more than 1, it is marked as a motion block; otherwise it is marked as a background block. Figure 7-2(c) shows the results of block-based mask with a block size of 8×8 pixels.

The selection of the block size is important because a large block may lead to a poor accuracy of motion estimation while a small block will result in many boundaries that cause artefacts during the SR process [77]. The selection of block size in the proposed method follows a divide-merge methodology. A pre-set small block size is applied to divide the image initially, and then the blocks which are adjacent and have the same property of motion characteristics are merged to reduce the boundary perimeter.

7.3.3 Motion Decomposition Module

This module aims to merge and transform the adjacent background blocks into background regions and segment the motion blocks into two types of blocks that have different motion characteristics.

The first step in this module is the extraction of background regions. This chapter introduces a unique background extraction method that aims to extract the background *regions* rather than *blocks* to reduce the boundary artefacts produced during the SR process. The inputs to the proposed background extraction algorithm are four consecutive LR frames, where the 2nd frame is chosen as the reference frame. All the adjacent background blocks are merged to reduce the boundary perimeter. Figure 7-3(a) shows the background blocks before the merging, where the red colour indicates the pixels on the boundaries. After the merging, the background now is represented by regions rather than blocks, as illustrated by Figure 7-3(b). By this mean, the perimeter of the background boundary is substantially reduced. In this example, the boundary pixels of background are reduced from 3840 pixels to 1120 pixels with a 71% reduction. To further reduce this type of artefacts, this chapter proposes to use only the background regions of the LR images for the registration process. This is achieved by taking the inner region of the LR frames outside the motion mask. The purpose of this step is to produce more accurate registration of the background regions, where the motion parts are removed and only the background regions are considered. It should be noted that, the reconstruction process is applied to the whole LR images. The original LR images are used for the reconstruction process based on the motion parameters from the background regions, by which means the artefacts around the boundary between background and motion regions can be significantly reduced.

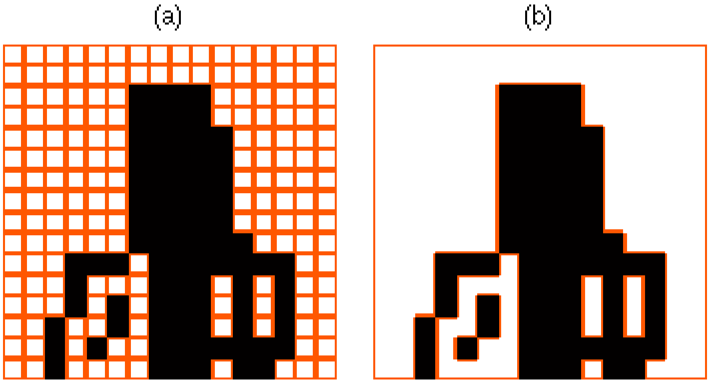


Figure 7-3 Results of detected background regions in the motion decomposition module, where (a) shows the background blocks (white) before applying this module and (b) shows the background regions after applying this module. The red colour indicates the boundary pixels for background.

The second step in this module is to decompose the motion blocks into two types: simple-motion blocks and complex-motion blocks. For the complex-motion blocks, due to the high complexity of motion objects, such as occlusion which refers to sudden disappearing of the existing objects and appearing of new objects, image registration methods may produce incorrect motion parameters that can substantially degrade the quality of super-resolved image. A single resolution enhancement method based on the reference frame only is therefore required for this type of blocks. For the simple-motion blocks, as the simple motion contains small motion and large motion, initially the simple-motion blocks were divided into small and large motion blocks and it was found that the Keren’s method works better for both types of blocks than other image registration methods, such as the Vandewalle's method. The rationale behind this process is that the Keren's method is suitable to estimate small motion and even large motion due to a coarse-to-fine image strategy used in this method, called a Gaussian pyramidal scheme. This pyramidal scheme uses different filtered, down-sampled versions of the original images to increase the precision for estimating large motion parameters. Therefore, even large motion will be transformed into small motions after this process [94], [146]. Hence, it is proposed to use simple-motion blocks rather than small and large-motion blocks in this chapter to reduce the computational cost.

The Keren's method uses the planar motion model based on three parameters to estimate the shifts and rotation parameters for each block of each LR video frame based on Taylor series expansions, including horizontal shift a_i, vertical shift b_i, and rotation angle θ_i $(i = 1,2,3,4)$. Euclidean distance of shifts is calculated based on

$$d_i = \sqrt{{a_i}^2 + {b_i}^2} \tag{7-9}$$

A preset threshold τ_1 is introduced to define the motion type of the considered block based on the below equation

$$\mathrm{MT} = \begin{cases} \text{simple} & \max\limits_{1\le i\le 4} d_i \le \tau_1 \\ \text{complex} & \max\limits_{1\le i\le 4} d_i > \tau_1 \end{cases} \tag{7-10}$$

The discussion for the selection of this parameter is presented in the next section. Figure 7-4 shows an example of motion decomposition module, where the left graph maps the maximum value of d_i and the right graph shows the visualisation of different types of motion blocks, where the three types of motion are clearly presented. Note that this example is for demonstration purpose only where the parameter is not optimally selected.

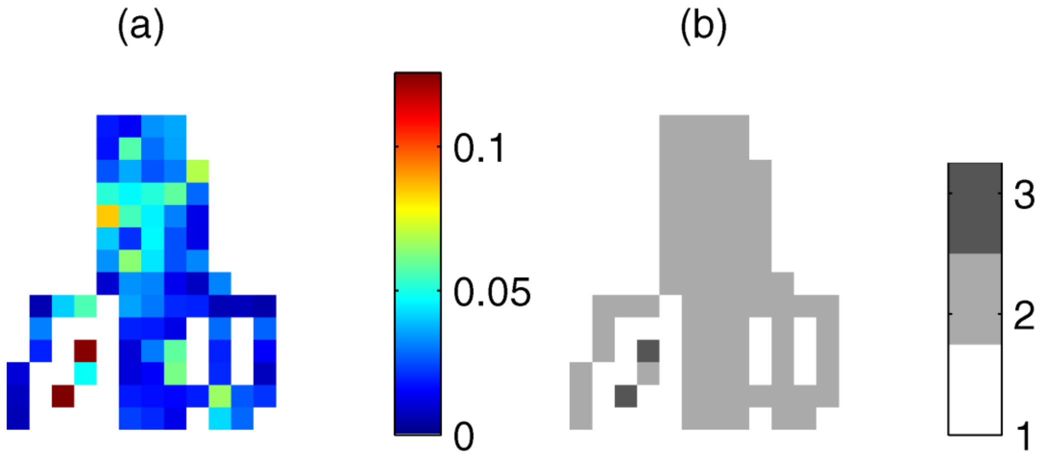


Figure 7-4 Results produced from the motion decomposition module, where (a) shows the maximum of d_i and (b) shows the divided three types of motion blocks/regions (white: background regions; light-gray: simple-motion blocks; dark-gray: complex-motion blocks).

To reduce the artefacts on the boundary of super-resolved motion blocks, each motion block is extended by α pixels at each side before applying the third module where α is the scale factor, which means that the block size becomes $(\mathrm{n}+\alpha)\ \times(\mathrm{n}+\alpha)$. The SR method will then be applied to the extended block, and α at each side of the HR blocks will be removed before merging in the reconstruction module.

7.3.4 Reconstruction Module

This module aims to super-resolve each motion block and background region and then merge them to reconstruct the final super-resolved frame.

For the complex motion blocks, a single wavelet-based resolution enhancement approach that uses only the reference frame is applied for this type of blocks. This chapter proposes to use a combination of WZP method with the scale factor $\alpha/2$ to produce the estimated LL sub-band and DWT-NEDI approach with the scale factor α to produce the estimated high-frequency sub-bands. The block diagram of this method is illustrated by Figure 7-5. The input is the LR motion block rather than the whole frame. The procedure can be summarised as:

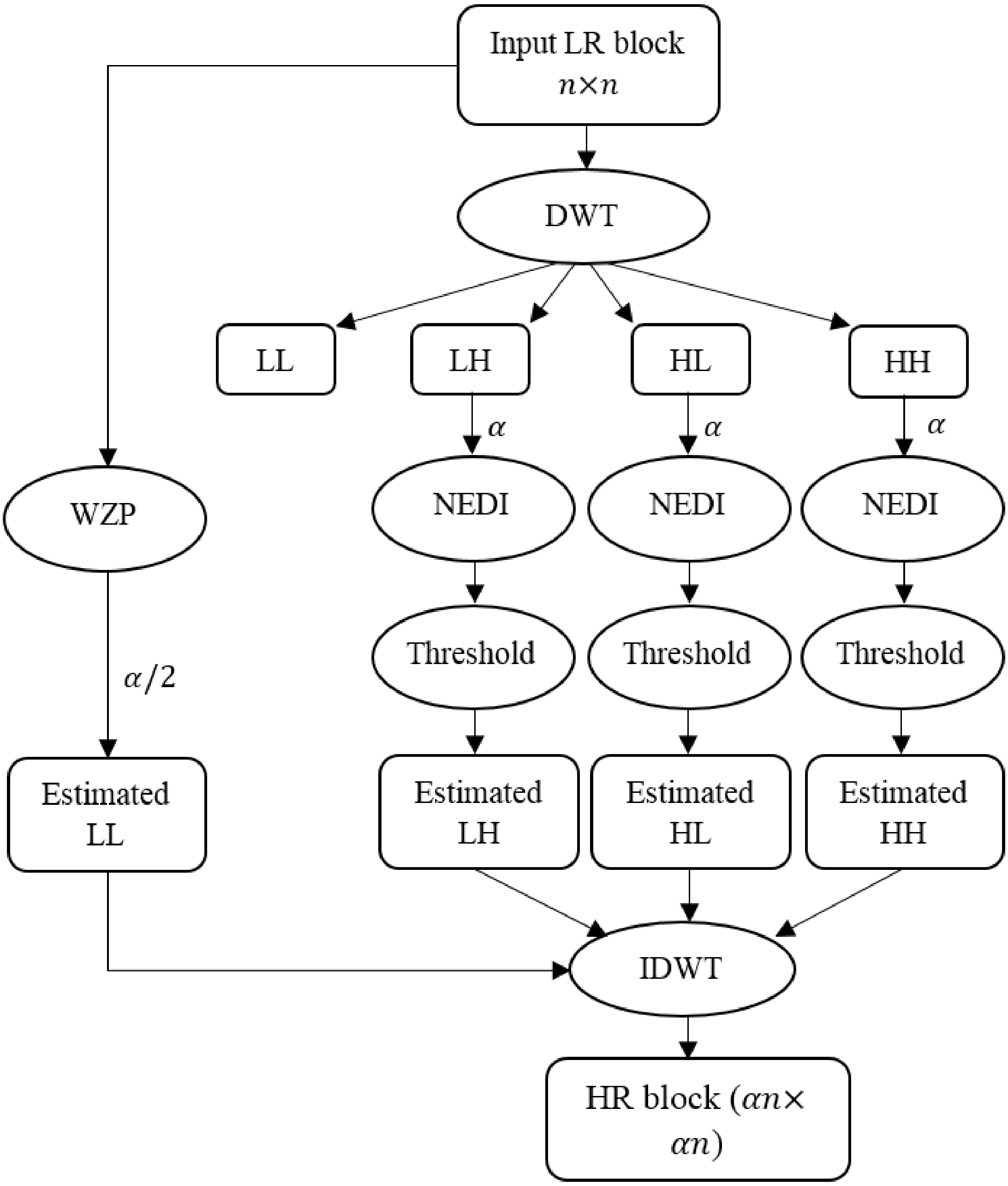


Figure 7-5 Block diagram of the proposed DWT-NEDI-WZP resolution enhancement approach for complex-motion blocks.

1 Apply one-level DWT decomposition of the input LR block to produce four frequency sub-bands (LL, LH, HL, HH);

2 Apply the NEDI method [87] to the high-frequency sub-bands (LH, HL, HH) with the scale factor α;

3 Apply an adaptive soft-thresholding process [141] for each high-frequency sub-band based on Eq. (7-11) to produce the estimated $\widehat{\mathrm{LH}}$, $\widehat{\mathrm{HL}}$ and $\widehat{\mathrm{HH}}$.

$$\widehat{\mathrm{LH}}(\mathrm{x,y}) = \begin{cases} \mathrm{LH}(\mathrm{x,y}) - \tau_d & \mathrm{LH}(\mathrm{x,y}) > \tau_d \\ 0 & |\mathrm{LH}(\mathrm{x,y})| \leq \tau_d \\ \mathrm{LH}(\mathrm{x,y}) + \tau_d & \mathrm{LH}(\mathrm{x,y}) < -\tau_d \end{cases} \tag{7-11}$$

The threshold τ_d can be calculated through

$$\tau_d = \sigma\sqrt{2\log(N)\,/N} \tag{7-12}$$

where σ is the standard deviation of the sub-band and N is the total number of pixels.

4 Apply WZP method [62] to the input LR block with the scale factor $\alpha/2$ to create the estimated $\widehat{\mathrm{LL}}$;

5 Apply the invert DWT (IDWT) using $\left(\widehat{\mathrm{LL}}, \widehat{\mathrm{LH}}, \widehat{\mathrm{HL}}, \widehat{\mathrm{HH}}\right)$ to produce the HR block;

A few DWT-based interpolation methods [70,71,72] have been developed to preserve the high-frequency components in the interpolated sub-bands. Nevertheless, the blurring effect from the employed interpolation method causes the potential loss of edges in these sub-bands. The combination of DWT with NEDI, namely DWT-NEDI, aims to improve the edge details of directional high-frequency sub-bands and reduce the annoying inter-directional interference in the SR process. This merit cannot be achieved using only the NEDI method. The application of this soft-thresholding function is based on the hypothesis that the large coefficients in the high-frequency sub-bands reflect the true edges of objects while the small coefficients reflect the noise.

For the background regions and simple-motion blocks, the Keren’s method is used to estimate the motion between the reference LR frame and its adjacent LR frames and then the structure-adaptive normalised convolution SANC [29] method with the scale factor $\alpha/2$ is used to produce the estimated LL sub-band.

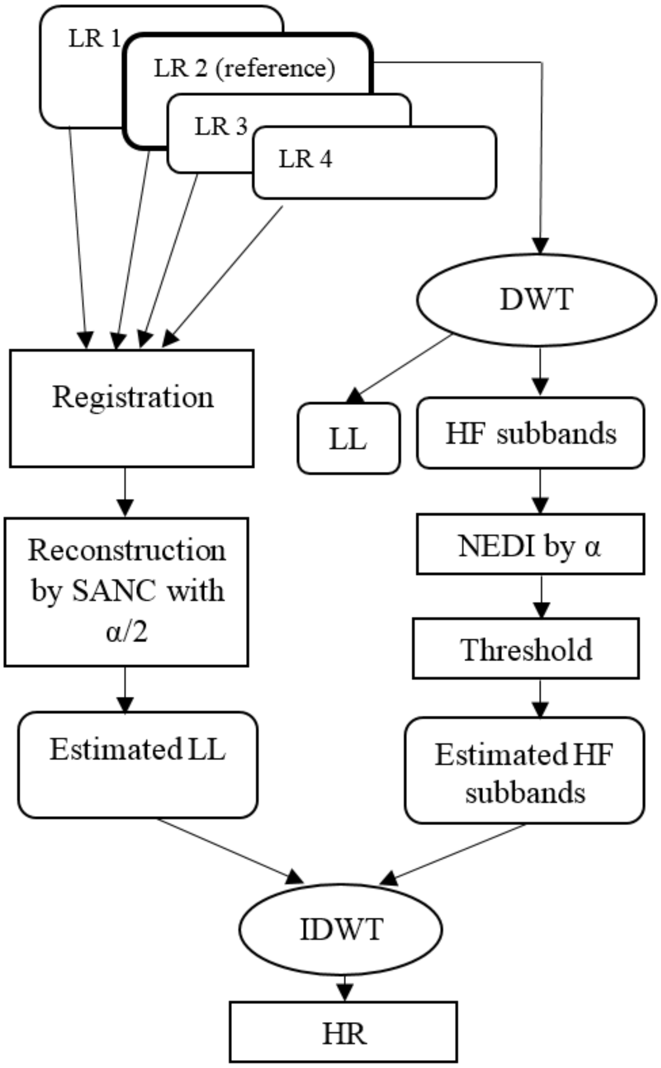


Figure 7-6 Block diagram of the proposed SR technique for simple -motion blocks and background regions.

The estimated high-frequency sub-bands are produced using the same procedure as shown in Figure 7-5 based on the reference frame only. The block diagram can be summarised by Figure 7-6.

Before merging all the HR regions and blocks together, the motion mask is used to select the HR background regions, and the boundary extension for motion blocks (α for each side) is removed.

7.4 Results and Discussions

In this section, four benchmark video sequences, namely, *"Akiyo", "Mother* & *daughter*", *"Foreman*" and *"Ice*" were tested to evaluate the performance of the proposed super-resolution technique. The video sequences were downloaded from a public database Xiph.org. The proposed algorithm and other methods for comparison were implemented using Matlab 2015. The original high-resolution test video sequences were resized to 512×512 pixels which are considered as the ground truth for performance evaluation. The reason to resize the original test videos to 512×512 pixels is for the convenience to evaluate the performance of the proposed SR technique in comparison with the other state-of-the-art resolution enhancement and SR reconstruction techniques in the literature. Based on the observation model, the input low-resolution video frames with the size of 128×128 pixels were produced from blurring and down-sampling each original HR video frame by applying twice cascade DWT with the db.9/7 wavelet function. The frame rate of the test videos is 30 frames per second and each of the videos has 300 frames. The scale factor was chosen as 4. The wavelet function db.9/7 was used in DWT for reconstruction.

7.4.1 Visual and Quantitative Performance Evaluation

Thirty frames were used to produce the motion mask, and the results of motion decomposition for the first 30 frames of *Akiyo, Mother, Foreman* and *Ice* are shown in Figure 7-7. It can be observed that *Akiyo* has only the simplest motion blocks, the motion of objects for *Mother* is dominated by simple-motion blocks and has very few complex-motion blocks while *Foreman* and *Ice* have significant complex-motion blocks, respectively.

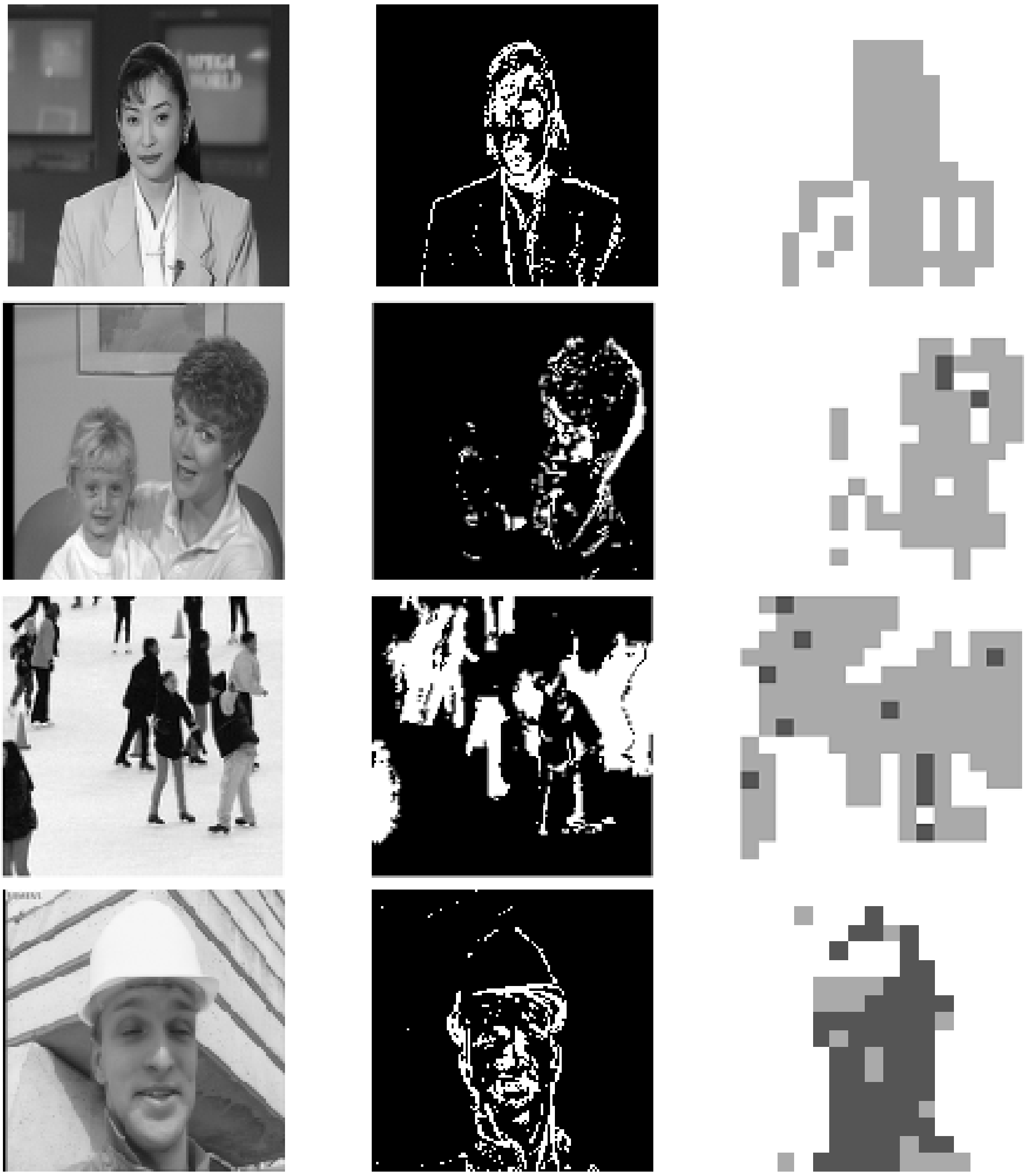


Figure 7-7 Results of motion decomposition for *Akiyo*, *Mother & daughter*, *Foreman* and *Ice*, respectively. The left column is the observed LR images, the middle column is the detected motion masks and the right column is the segmented three motion types. The mapping between colour and motion type can be found in Figure 7-4.

Figure 7-8 shows the randomly selected super-resolved frames using the proposed method comparing with other methods for a selected region. It can be clearly observed that the proposed algorithm is more efficient for representing the required complex areas of the original HR video frame.

Figure 7-8 Visual results of HR images using the proposed method in comparison with other methods. The first row is the 27th frame of the *Akiyo*; the second row is the 49th frame of the *Mother & daughter*; the third row is the 2nd frame of the *Foreman*, and the fourth row is the 19th frame of the *Ice*. The first column is the observed LR images; the second column is the interpolated HR images using bicubic; the third column is the interpolated HR images using NEDI and the fourth column is the super-resolved images using the proposed ALMD-WTR method.

For example, the lips in *Foreman's* mouth and the teeth in *Mother's* mouth are better represented by the proposed SR algorithm, whereas these areas are not described well by the resolution enhanced methods (NEDI and Bicubic). Additionally, the proposed technique preserves the edges of the original HR video frame without burring. For example, the edges of the face in *Foreman and Akiyo*, and the ice skates of the top right person in *Ice* produced by the proposed technique are much distinct in comparison to the resolution enhanced images produced by NEDI and Bicubic. Moreover, the aliasing artefacts are reduced by the proposed method in comparison to other methods, for example, the shoulders of *Akiyo, Mother*, and the stripe of the wall in *Foreman.* It can be concluded from the above observations that the proposed algorithm produces the best visual quality among the considered methods in terms of representing the complex areas, preserving the edges and reducing the aliasing artefacts.

The proposed method uses WZP method to estimate the LL sub-band for the complex-motion blocks, as illustrated by Figure 7-5. Empirical tests have been conducted using other two resolution enhancement methods: bicubic and NEDI. As an example, the results of the 4nd frame of *Foreman* are shown in Figure 7-9. Inspection of Figure 7-9(b)-(d), shows that the PSNR gain between the proposed method using bicubic, NEDI, and WZP and the nearest neighbour interpolation. It clearly proves that the proposed WZP method has the most improvement with less influence on the boundary, where the face area has the most significant improvement.

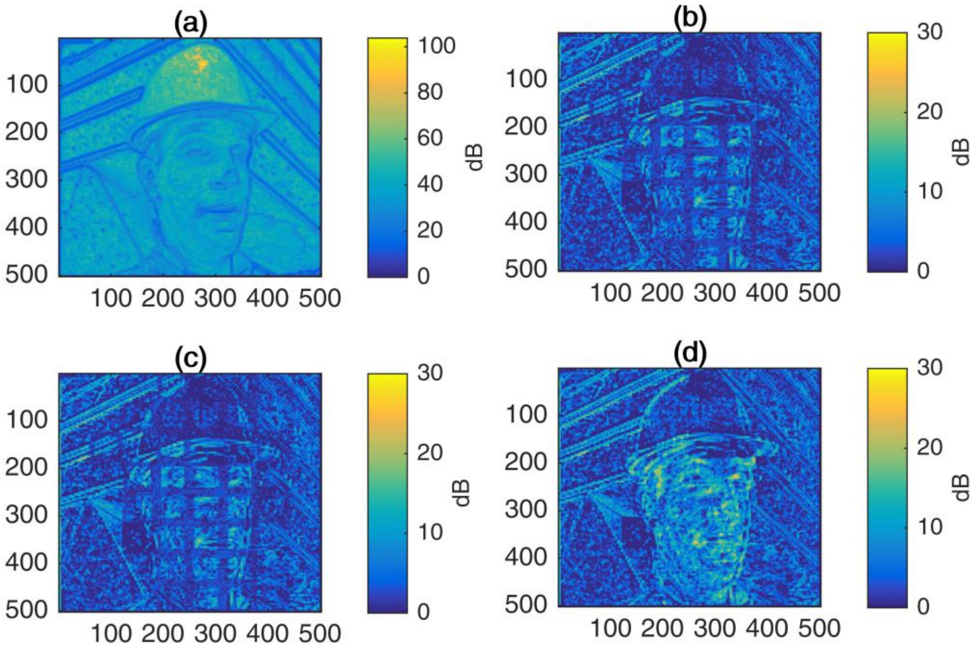


Figure 7-9 An example to show the results using different methods to estimate the LL sub-band for the complex blocks. (a) The distribution of PSNR between the nearest neighbour interpolation and the raw image; (b) the PSNR gain between the proposed method using bicubic to estimate LL and the nearest neighbour; (c) the PSNR gain between the proposed method using NEDI to estimate LL and the nearest neighbour; (d) the PSNR gain between the proposed method using WZP to estimate LL and the nearest neighbour.

To further investigate the improvement of the proposed method, Figure 7-10 shows the local PSNR maps for different scenarios for the example of *Mother*. The local PSNR map was calculated by a 5×5 pixels window. Figure 7-10(a) shows the PSNR distribution between the raw HR image and the interpolated HR image using the nearest neighbour method, which indicates the location of noise introduced by the degrading process. Regions of the human body have more information loss (blue regions) introduced by the degrading process, while background regions have less information loss (yellow regions). The blue regions tend to be areas with fine features (like boundary of the human body or the photo frame), while the yellow regions tend to have more coarse structures. Similar representation using the proposed method is shown in Figure 7-10(b), inspection of which proves that background has been improved slightly while fine boundaries have been much better improved. To break down the contribution of each component, Figure 7-10(c) and (d) show the PSNR gain of the proposed method over bicubic and NEDI, respectively. It has been observed that fine features are significantly improved in comparison to bicubic due to the consideration of adjacent frames, while the improvement of coarse features is relatively small. Fine features are further improved in comparison to NEDI while the improvement of coarse structures is very small. All these observations clearly demonstrate that the proposed method improves the quality of both background and true edges with different levels.

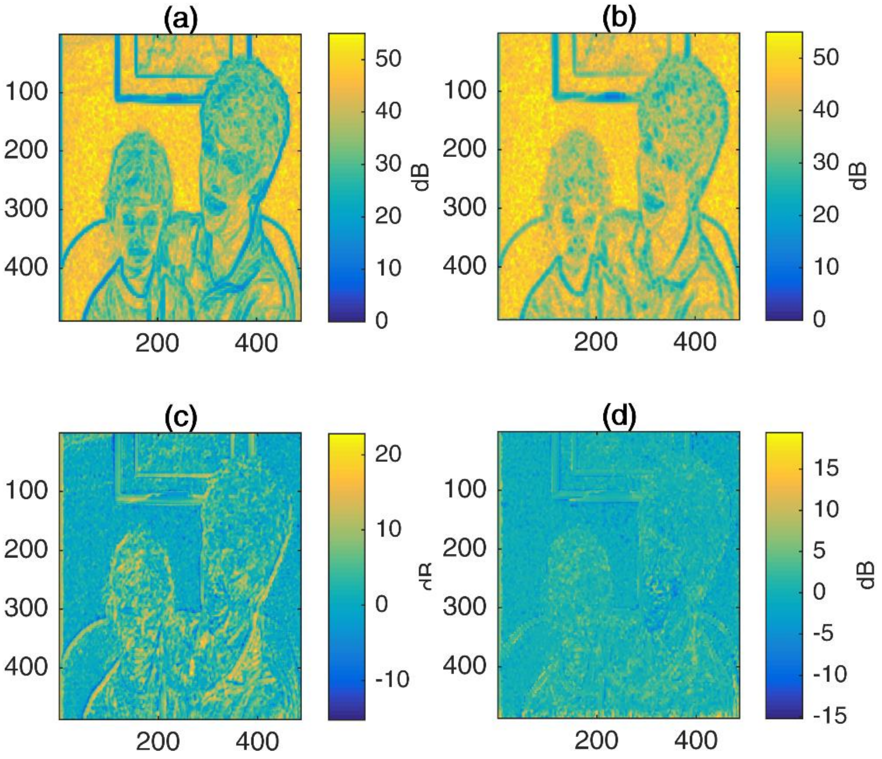


Figure 7-10 An example to show the improved PSNR distribution of the proposed technique. (a) The distribution of PSNR between the interpolated image using the nearest neighbour and the raw image; (b) the distribution of PSNR between the super-resolved image using the proposed method and the raw image; (c) the PSNR gain between the proposed method and bicubic interpolation; (d) the PSNR gain between the proposed method and NEDI.

The improvement of the proposed method could be better demonstrated by using the modified metric, PSNR-HVS-M [147] based on properties of a human visual system (HVS). This metric can also be calculated for each image block as the local PSNR map. However, the advantage of PSNR-HVS-M measure is that it takes into account human perception, while the standard PSNR doesn’t relate to human perception.

Table 7-1 shows the comparison of the average PSNR and SSIM values of the 100 frames from the proposed method and other interpolation methods for the four test videos, where the motion block size was chosen as 32×32. For achieving statistical significance, the first 100 frames of each video were tested. Experiments show that there is no big difference in results if more frames are considered. It can be observed that the proposed method consistently produces the highest PSNR and SSIM values for each video sequence. This is because that NEDI, bicubic, and nearest methods use only spatial information in the reference frame, while the proposed method uses spatio-temporal information of the reference frame and other neighbouring frames. Additionally, DWT-based SR reconstruction approach is more effective to recover the missing high-frequency details of the given LR video frames, where the true edges are preserved and the noise is removed benefiting from the nonlinear soft-thresholding.

Table 7-1 The average PSNR and SSIM values of the 100 frames by the proposed technique in comparison with other interpolation methods, where the block size is 32 by 32

Methods	PSNR				SSIM			
	Akiyo	*Mother*	*Foreman*	*Ice*	*Akiyo*	*Mother*	*Foreman*	*Ice*
Nearest	26.18	28.05	25.30	25.27	0.37	0.31	0.38	0.29
Bicubic	27.33	29.24	26.84	26.58	0.45	0.37	0.47	0.34
NEDI	30.47	32.43	28.34	30.08	**0.53**	0.47	0.53	**0.44**
Proposed method	**31.51**	**33.57**	**31.60**	**31.43**	**0.53**	**0.48**	**0.58**	0.40

Table 7-2 The average PSNR values by the proposed technique in comparison with other state-of-the-art methods, where the block size is 32 by 32

Methods	*Foreman*
Protter et al. [112]	29.01
Izadpanahi et al. [77]	31.51
Proposed technique	**31.60**

Furthermore, the PSNR increment percentage of the proposed method over the NEDI method is 3.4%, 3.5% and 11.5% respectively. The proposed method was also compared with state-of-the-art SR methods [77,112], and the results for one of the examples with large proportion of complex-motion blocks (*Foreman*) are shown in Table 7-2, which clearly demonstrates that the proposed technique has the best performance for objects with complex motions. As shown in Table 7-2, the PSNR increment percentage of the proposed method is 8.9% over Protter et al. [112] and the increment is 0.3% over Izadpanahi et al. [77], which illustrates a superior performance against the compared state-of-the art methods.

7.4.2 Parameters Selection

The effectiveness of the size of motion block on the performance of the proposed technique is firstly discussed. Based on object characteristics, the four test video sequences are classified into two types: simple-motion objects (*Akiyo* and *Mother*), and complex-motion objects (*Foreman* and *Ice*). Table 7-3 shows the average PSNR and SSIM values of the 100 frames produced by the proposed technique with different motion block sizes on the four test video sequences. It has been observed that the size of block can affect the performance significantly, especially when the motion of objects is complex. For *Akiyo* and *Mother* with simple-motion objects, the results are almost the same (with a standard deviation of PSNR of 0.09 dB and 0.21 dB respectively). This may be due to the contribution of motion compensation is very limited. For *Foreman* and *Ice* with complex-motion objects, the block size has more influence on the results (with a standard deviation of 0.74 dB and 0.40 dB respectively). It has been also observed that a relatively large block sizes produce better results for videos with complex-motion objects, while for videos with simple-motion objects, small block sizes produce better results. Collectively, a 32×32 block size is suggested for the proposed method.

Table 7-3 The PSNR and SSIM values of the 100 frames for four examples produced by the proposed technique using different motion block sizes

Block size	PSNR(SSIM)			
	Akiyo	*Mother*	*Foreman*	*Ice*
8×8	**31.67(0.54)**	**33.68(0.48)**	30.80(0.55)	30.51(0.38)
16×16	31.62(0.53)	33.64(0.48)	31.54(0.57)	30.82(0.39)
32×32	31.51(0.53)	33.57(0.48)	**31.60(0.58)**	**31.43(0.40)**
64×64	31.47(0.53)	33.21(0.46)	30.02(0.54)	31.17(0.40)

Figure 7-11 Motion detection results using (a) a small threshold, (b) the Otsu's method, and (c) a large threshold

The proposed technique also requires several other pre-set parameters. In the motion detection module, the selection of parameter β which represents the updating rate has been recommended as 0.05. The selection of threshold τ_p which determines the quality of motion distribution has been discussed in many experiments. Figure 7-11 shows motion detection results using the Otsu's method, a small and large values of τ_p. It is clearly shown that the proposed method using the Otsu's method Figure 7-11(b) produces a more reasonable motion distribution in comparison to those from a too large and small threshold. For Figure 7-11(a), motions are detected on the wall of the background, which will result in many motion blocks on this background region. Although the same registration method is applied in this region, no matter the selection of threshold, the artefacts around the motion blocks will degrade the image quality. For Figure 7-11(c), some small and moderate motions are treated as background, which will result in inaccurate local motion estimation.

Table 7-4 The PSNR values for the second frame with different values of τ_1, where the block size is 32 by 32

τ_1	*Akiyo*	*Mother*	*Foreman*	*Ice*
0.2	31.51	32.61	31.22	31.13
0.5	31.51	33.57	31.60	31.43
1	31.51	33.57	31.60	31.43

In the motion decomposition module, experiments show that when the parameter τ_1 is selected between 0.1 to 1, the proposed method performs well. Table 7-4 shows the PSNR values when τ_1 is chosen as 0.2, 0.5 and 1 respectively. It is observed that the variation of result is not significant in the considered range. For *Akiyo,* since all values of d_i are smaller than 0.2, the results are identical. For all tested videos, the selection of 0.5 is recommended.

The selection of threshold τ_d in the reconstruction module is determined by Eq. (7-12). This parameter controls the balance between noise removal and edge preservation.

7.4.3 Performance from Different Motion Detection Methods

This sub-section is dedicated to show the performance of the proposed technique using different motion-based background subtraction methods. Two state-of-the-art background subtraction methods, including simple background subtraction SBS and sigma delta estimation SDE, have been implemented on the four test videos. The average PSNR and SSIM results of the first 10 video frames are shown in Table 7-5. It has been observed that the PSNR and SSIM values produced by the proposed method of each video sequence are relatively similar.

Table 7-5 The average PSNR and SSIM values of the first 10 frames produced by the proposed technique using different motion-based background subtraction methods, where the block size is 32 by 32

Motion detection methods	PSNR(SSIM)			
	Akiyo	*Mother*	*Foreman*	*Ice*
SBS	31.45(0.53)	33.53(0.45)	32.02(0.6)	30.51(0.45)
SDE	31.42(0.53)	33.52(0.45)	32.13(0.6)	30.48(0.44)
RA	31.42(0.53)	33.53(0.45)	32.13(0.6)	30.49(0.44)

7.4.4 Performance against Noise

To evaluate the performance of the proposed method against noise in the LR images, three different levels of noise, 35dB, 25 dB and 20dB respectively, have been applied into the 2nd frame of the *Mother* video. Ten tests for each noise level were conducted and the results were averaged. Results including PSNR and SSIM are shown in Table 7-6. It has been observed that the proposed technique produces the highest PSNR and SSIM values for all three cases among the selected methods, which demonstrates the good robustness of ALMD-WTR against noise. It also has been observed that the merit of the proposed technique against higher level of noise is more significant.

Table 7-6 The average PSNR and SSIM values of the second frames produced by the proposed technique with different levels of noise for the Mother video, where 10 tests were repeated

Methods	PSNR			SSIM		
	35dB	25dB	20 dB	35dB	25dB	20 dB
Nearest	27.85	26.39	24.03	0.24	0.16	0.11
Bicubic	29.07	27.79	25.56	0.31	0.22	0.16
NEDI	32.21	30.47	27.94	0.41	0.30	0.23
Proposed	**33.40**	**31.63**	**29.65**	**0.44**	**0.33**	**0.26**

For example, when the SNR is reduced from 35 dB to 20 dB, the PSNR value drops 13.7%, 12.1%, 13.3% and 12.46% for nearest, bicubic, NEDI and ALMD-WTR respectively. The SSIM value drops 54.2%, 48.4%, 43.9% and 40.9% respectively. These observations are a clear evidence that the proposed technique can reduce noise benefiting from the soft-thresholding process.

7.5 Summary

This chapter proposes a new video super-resolution reconstruction technique, called ALMD-WTR based on merits from both the frequency and spatial-domains, especially for the real-world videos with complex local motions between video frames. The introduced method has the following features:

- The frame is divided into three types of blocks/regions based on motion characteristics, and each type is processed with a tailored registration and reconstruction methods or a single resolution enhancement method. For the complex-motion blocks, a single wavelet-based image resolution enhancement method based on WZP, and DWT-NEDI is introduced without considering the adjacent frames. For the simple-motion blocks and background regions, to achieve more accurate motion estimation, the Keren’s registration method is used to estimate the motion between the reference frame and its adjacent frames.
- The combination of DWT and NEDI is unique. It is able to improve the edge details of directional high-frequency sub-bands and hence decrease the undesirable inter-directional interference in the SR process. A nonlinear soft-thresholding process on the high-frequency sub-bands produced by DWT is proposed to remove the noise and preserve the true edges.
- The artefacts caused by the boundary have been reduced by three ways: (a) the adjacent background blocks are connected and transformed into background regions, by which means the perimeter of background boundary is significantly reduced; (b) the reconstruction process for the background regions is applied to the whole image rather than only the background regions; (c) each motion block is extended by α pixels at

each side for registration and reconstruction processes, and the scaled extension is then removed before merging all the HR blocks/regions.

The performance of the proposed method has been evaluated through testing four benchmark video sequences in comparison with other interpolation methods and state-of-the-art SR techniques. Both subjective and objective results show that the proposed approach outperforms the compared methods. It has been found that the relatively small-motion block sizes perform better than the large-motion block sizes for all four tested videos. Although, the small-motion block sizes cause artefacts around the boundary, the proposed method has successfully suppressed more artefacts around boundaries. It has also been found that for complex-motion objects, the selection of block size has much influence on the results than that for simple-motion objects.

A limitation of this technique is that it requires a number of pre-set parameters. An optimisation of these parameters requires further studies in the future.

8 Conclusions and future work

This chapter describes the fulfilment of the research aim and objectives, explains the key research challenges, summarises the research contributions, discusses the main research limitations, highlights the most significant conclusions, and finally suggests recommendations for future work in the wavelet-based super-resolution research area. The chapter is divided into 5 sections: Section 8.1 details the accomplishment of the research aim and objectives, along with the research challenges; Section 8.2 describes the research contributions; Section 8.3 discusses the research limitations; Section 8.4 draws the central conclusions; and Section 8.5 suggests recommendations for future work.

8.1 Fulfilment of aim and objectives (main challenges)

This section discusses the challenges and subsequent fulfilment of the four research objectives outlined in Chapter 1. A description for each objective and corresponding achievement is given below:

(1) Design a novel performance assessment approach for improving the wavelet-based image resolution enhancement techniques.

In order to achieve this objective, a comprehensive literature review, as well as a taxonomy of the existing wavelet-based image resolution enhancement techniques were implemented, and a summary of these techniques in terms of the way to evaluate their performance was conducted. It was observed that the existing methods make a number of assumptions with regard to the factors that affect the performance of the wavelet-based methods, such as the method of producing the observed LR images and the selection of wavelet functions. This limits the performance of these methods in practical applications. It also was observed that there are inconsistencies in the assumptions regarding the factors considered for each individual method. And for some methods, the factors are even not described. All these observations inspired this project to identify the factors that effectively influence on the performance of these methods, and to then quantitatively assess the impact of the prevailing assumptions.

Notwithstanding the fact that the performance of the resolution enhancement techniques under consideration were assessed to analyse the importance of each factor by varying one factor and fixing others, another challenge of this objective was how to improve the best developed techniques with optimal factor selection. Thus, this project proposed a novel optimal factor analysis (OFA) approach to improve the performance of the analysed wavelet-based image resolution enhancement techniques, and assess the performance of these techniques in a more comprehensive way: by selecting the best technique and simultaneously analysing the optimal factors. A new figure of merit measure was also introduced in this project to better evaluate the overall performance.

(2) Develop a new wavelet-based image resolution enhancement method for increasing the spatial resolution of a single LR image.

In order to achieve this objective, the issues, assumptions, important factors, and methods relative to wavelet-based image resolution enhancement were revealed and studied. Although many state-of-the-art wavelet-based methods have been introduced in different fields in order to reconstruct the high-resolution HR image from an observed LR image, resolution enhancement of satellite imaging is in high demanded for a number of applications, particularly for remote sensing. This is because remote sensing image data can have errors in geometry, and analysis from these low-quality (degraded) images can be extremely difficult.

Ultimately, success was achieved by developing a new image resolution enhancement method based on DWT and NEDI (DWT-NEDI) for degraded satellite images; this was to compensate for the geometric errors and improve the edge details of high-frequency sub-bands. An adaptive threshold is applied to boost the edges and eliminate potential noise in the estimated high-frequency sub-bands. This approach is based on using the interpolation of isolated high-frequency sub-bands produced by DWT in order to preserve better edges of the image, rather than using a direct interpolation technique which blurs the edges and produces artefacts.

The DWT-NEDI algorithm was tested on 20 different satellite images divided into five groups, so as to evaluate the variation in performance in terms of different types of images. The performance of the DWT-NEDI method was also evaluated in comparison with the conventional interpolation and state-of-the-art wavelet-based image resolution enhancement methods. Furthermore, four new criteria were proposed to better assess the overall performance for a number of images.

(3) Develop a robust wavelet-based multi-frame (video) super-resolution method for increasing the spatial resolution when the camera is moving and the observed object is static.

This objective was achieved through developing a robust wavelet-based super-resolution SR technique based on a combination of DWT-NEDI and a nonlinear soft-thresholding, with reconstruction-based SR techniques for noisy LR video sequences containing global motion to increase the spatial resolution and recover the noiseless high-frequency details. The technique developed is especially useful in particular situations when the camera is moving and the observed scene is stationary. The algorithm proposed is based on applying a planar motion estimation, by the Keren method, to estimate the motion parameters between the reference LR frame and its neighbouring LR frames, and then applying the structure adaptive normalised convolution, SANC method, to combine the registered frames and produce the estimated low-frequency sub-band.

Although most of the existing wavelet-based SR methods also assume global motion between video frames, these methods have limited performance capabilities for a variety of noise levels, motion levels, wavelet functions, and number of LR frames used. Therefore, another challenge of this objective is to develop a robust SR method that can provide flexibility for these various factors. To address this, the proposed algorithm was modified and it was analysed how variety of these factors can affect the performance, and the effectiveness of each factor was also discussed. The proposed robust SR method was tested on three-well known video sequences (a total 100 frames for each) and the results were compared with conventional interpolation, state-of-the art resolution enhancement, and classic SR methods.

(4) Design a new wavelet-based multi-frame (video) super-resolution framework for increasing the spatial resolution when the camera is stationary and the observed object is moving.

In order to achieve this objective, a new wavelet-based SR framework, called adaptive local motion decomposition and wavelet transform reconstruction (ALMD-WTR), is proposed in order to address the challenge of the SR problem for real-wold video sequences containing complex local motion.

The new framework proposed to solve the problem through three modules: motion detection module, motion decomposition module, and reconstruction module. The aim of the motion detection module was to divide the observed LR video frames into background blocks and motion blocks. This module was achieved by introducing the running average (RA) method to detect moving objects from an adaptive background model. A binary pixel-based motion mask, based on a threshold, was then produced; where the Otsus method was applied to choose the optimal threshold. To construct the units for processing, it was proposed to transform the pixel-based mask into a block-based mask for differentiating the motion blocks from the background blocks.

There are two steps in the motion decomposition module. The first step was to transform the adjacent background blocks into background regions and extract the background regions for reduction of background boundary artefacts. In order to achieve this process, the researcher proposed a unique background extraction algorithm to extract the background regions rather than blocks. The second step was to decompose the motion blocks into two types of blocks: simple-motion blocks and complex-motion blocks.

The reconstruction module aimed at producing HR blocks for the complex-motion blocks, simple-motion blocks and background regions, and then combining them to reconstruct the super-resolved frame. When producing HR complex-motion blocks, as the motions of the objects can be very complex, image registration algorithms may produce incorrect motion vectors. Therefore, a resolution enhancement method is required for this type of blocks, based on consideration of the reference frame only. Following on from the research conducted in objective two, the researcher proposed a unique resolution enhancement approach by combining the DWT-NEDI method and WZP method to produce the estimated high-frequency and low-frequency sub-bands respectively, and then combining them to reconstruct the HR blocks. For the simple-motion blocks, the author proposed dividing the simple-motion blocks into small and large-motion blocks due to the simple motion contains small and large motion. Based on the rationale that the Keren method performs well for both types of blocks, the author proposed using simple-motion blocks, instead of large and small-motion blocks, in order to reduce the computational cost. Therefore, to produce the HR simple-motion blocks, the Keren method is used to estimate the motion for this type of blocks, and then the structure-adaptive normalised convolution SANC is used to produce the estimated low-frequency sub-band. On the other hand, the estimated high-frequency sub-bands are produced using the contributed DWT-NEDI algorithm.

8.2 Research contributions

The main contributions of this research are summarised below:

(1) **Designing of a novel optimal factor analysis (OFA) approach to improve the performance of the wavelet-based image resolution techniques.** This research analyses the important factors that affect the performance of the wavelet-based techniques and then reveals how to utilise these factors to assess and improve the performance, by the design of a novel OFA approach. The OFA algorithm assesses the performance of the selected methods in a more comprehensive and equitable way so as to increase the applicability and fidelity of these methods, as well as better assess their overall performance.

(2) **Developing of a new image resolution method using DWT and NEDI for application with degraded satellite images to correct the geometric distortion and remove potential noise.** This research develops DWT-NEDI algorithm based on integrating merits from both the frequency and spatial-domains for improving the spatial resolution and preserving more edges of directional high-frequency sub-bands. The novelty of this algorithm is the introduction and integration of a nonlinear soft thresholding process in order to better preserve the edges and remove the noise.

(3) **Developing of a robust global-based video SR method using discrete wavelet transform reconstruction to increase the spatial resolution.** This research contributes a robust video SR algorithm of the observed noisy LR video frames captured from a moving camera with global motion for increasing the spatial resolution whilst supressing the noise and aliasing artefacts. The algorithm is based on combining the Keren registration method, followed by SANC reconstruction method, with the DWT-NEDI approach. Another contribution of this research is that the algorithm proposed can be modified to provide flexibility with various motion levels, noise levels, wavelet functions, and number of used LR frames.

(4) **Designing of a new local-based video SR framework using adaptive local motion decomposition and wavelet transform reconstruction (ALMD-WTR) to increase the spatial resolution.** The new framework aims at addressing the challenge of practical video sequences with complex local motion between frames. This framework contributes to increasing image resolution through solving the local registration errors. This is achieved from: adaptive motion decomposition; reducing the boundary artefacts by reduction of background boundary; and extension of motion blocks.

8.3 Research limitations

Although, the research aim and objectives were successfully achieved, there were some limitations relating to the methods developed by this project, as well as those of other SR methods.

Firstly, there was a limitation of the mechanism to produce the observed LR images. This is because there is no validated model that can fully describe this underlying mechanism, and this mechanism is unclear, and also varies case by case due to the laws of physics.

Secondly, there were very limited knowledge to guide the selection of the number of LR images used to produce the HR image in SR reconstruction.

Thirdly, there was difficulty in building a practical SR system. This is because the observed LR images contains large amounts of aliasing artefacts. The performance of the registration algorithms degrades rapidly when the resolution of the observed LR images goes down.

Fourthly, there were difficulties in evaluating the performance of the proposed DWT-NEDI technique compared to other techniques for a number of the test images. This is because the available criteria are used for evaluating the performance of the considered techniques for a single image. This adds challenges in the validation process.

Finally, there were limitations in selection of the pre-set parameters and difficulties in selection of the optimal thresholds for the proposed ALMD-WTR method.

8.4 Discussion and Conclusions

This research addresses an important issue in SR reconstruction regarding the wavelet-based SR reconstruction problem. The following overall points are concluded:

- The wavelet-domain-based SR reconstruction problem is different from the spatial-domain or frequency-domain-based SR reconstruction problems in terms of the associated strategies, achieved challenges, and available approaches. The wavelet-domain-based SR exploits both the spatial and frequency-domains, and integrates the merits of both to solve the SR reconstruction problem. The wavelet transform provides both frequency information and temporal information in the transformation process.

- The wavelet transform is an effective tool that decomposes an image into low and high-frequency sub-bands, and then studies each sub-band with a resolution matched to its scale. The attractive multi-resolution property of wavelet transform enables it to analysis the image data at more than one resolution. The advantage of this strategy is that image characteristics can be isolated and examined such that global characteristics can be examined at coarse scales, while local characteristics can be analysed at fine scales.
- The wavelet-based SR problem is an advanced technique developed initially from the wavelet-based resolution enhancement problem. However, the wavelet-based SR is different from the wavelet-based resolution enhancement in terms of the number of LR images, the quality of reconstructed image, associated challenges, and existing algorithms. The wavelet-based SR aims to reconstruct one HR image, or a sequence of HR images, from multiple different LR images acquired from the same scene, while the wavelet-based resolution enhancement aims to reconstruct one HR image from one LR image.
- This research was divided into 4 research dependent objectives. Objectives 1 and 2 were based on addressing the wavelet-based resolution enhancement problem; Objectives 3 and 4 were based on addressing the wavelet-based SR problem. Objective 1 was to design the OFA approach for improving the performance of wavelet-based resolution enhancement methods. Objective 2 was to develop an improved resolution enhancement DWT-NEDI approach for improving the spatial resolution. Objective 3 was underpinned by Objective 2 from the combination of the DWT-NEDI method with SR methods for increasing the spatial resolution. Objective 4 was underpinned by

Objectives 2 and 3 for increasing the spatial resolution. The most significant conclusions of this research include:

- The factors that can substantially affect the performance of wavelet-based resolution enhancement methods include: the way of generating the LR image; the wavelet family and its wavelet function; and the scale factor. The factors that can moderately influence on the performance of these methods are the selection of the interpolation method, and the selection of the test image.
- The introduced OFA approach is able to improve the performance of the wavelet-based methods considered and increase the applicability and fidelity of these methods. The quantitative results show that the OFA method significantly improves the performance of the WZP method and has the potential to be extended to other wavelet-based methods. Notwithstanding the significant quantitative results, the visual results indicate that the difference between the selected methods can be small; and it is difficult to be inspected visually.
- The combination of DWT-NEDI is able to improve edges of directional high-frequency sub-bands and thus reduce the annoying inter-directional interference in the resolution enhancement process. An adaptive threshold on the high-frequency sub-bands is proposed to remove the noise and preserve the true edges.
- The DWT-NEDI approach contributed here is based on the interpolation of isolated high-frequency sub-bands produced by DWT, to preserve more edges than when using a direct interpolation by conventional interpolation methods, which produce blurred edges and undesirable artefacts. This benefit cannot be achieved using only DWT or NEDI.
- The visual results demonstrate the ability of the DWT-NEDI method to improve the observed LR images by providing more edges, potentially offering more details in selected regions. The DWT-NEDI algorithm can perform well with the db.9/7 wavelet function, and even better with other wavelet functions, although the difference between them is relatively small.

- The proposed global-based SR technique is based on combining the resolution enhancement DWT-NEDI approach with the Keren registration and SANC reconstruction methods for increasing the spatial resolution. One of the motivations for this approach is to deal with various motion levels, noise levels, wavelet functions, and the sufficient number of frames.
- The visual results indicate that the proposed SR technique can better preserve the edges and remove the noise and aliasing artefacts. The quantitative results show that the algorithm contributed here produces a (16% and 11% increment over Keren-SANC for the videos *Mother* and *Akiyo* respectively), and a (17% increment over Keren-SANC for the video *Foreman*).
- The global SR algorithm introduced is able to deal with various motion levels, noise levels, wavelet functions, and the requisite number of frames. Different motion levels can affect performance; the highest PSNR and SSIM values were produced when the motion level is relatively small; however, when the motion level is large, these values decrease. There was a 12% and 7% decrease of PSNR and SSIM values respectively, observed from the smallest to largest motions considered. Different number of used frames can affect performance; higher PSNR and SSIM were achieved with a higher number of sampled frames; however, only about a 1% increase of PSNR and SSIM values was observed when the number of frames considered changed from 4 to 32.
- The ALMD-WTR framework is able to decompose the observed LR frame into simple-motion blocks, complex-motion blocks, and background regions, and produce HR blocks/regions for each type using a tailored registration and reconstruction methods, or a wavelet-based resolution enhancement approach, and finally reconstruct a super-resolved frame by combining all of them.
- For the complex-motion blocks, a combination of WZP and DWT-NEDI approach, based on the reference frame only, is proposed for this type of

blocks, to produce the HR blocks. For the simple-motion blocks and background regions, the Keren method and SANC method are applied for this type of blocks/regions to produce the HR blocks and regions.

- The ALMD-WTR method aims to reduce the artefacts caused by background boundary through transforming the adjacent background blocks into background regions, applying the registration process on only the background regions to produce more accurate registration of the background regions in which the motion parts are removed and only the background regions are considered, while applying the reconstruction process on the whole image This method also aims to reduce the artefacts caused by motion blocks boundary through extension each motion block by the scale factor α pixels at each side for the registration and reconstruction processes, and the scaled extension is then removed before combining all the HR blocks.
- The selection of block size can significantly affect performance, especially when the motion of objects is complex. For videos *Akiyo* and *Mother,* with simple-motion objects, the block size has less influence on the performance (with a standard deviation of PSNR of 0.09 dB and 0.21 dB respectively) and a small block size produces better results. For videos *Foreman* and *Ice,* with complex-motion objects, the block size has more influence on the performance (with a standard deviation of 0.74 dB and 0.40 dB respectively) and a relatively large block size produces better results.

8.5 Future work

This PhD project focused on the study of the wavelet-based SR reconstruction problem, which could be beneficially explored further in the field of SR reconstruction. Recommended suggestions for future work include:

- Extending the proposed SR algorithms to address the wavelet-based SR reconstruction problem for increasing the spatial and temporal resolution for video sequences containing complex local motions. This research aims to recover very fast local motions that are not seen or acquired correctly in the video sequence. This future work will introduce new visual capabilities of very fast moving objects by handling together the two visual effects caused by motion aliasing and motion blur.

- Extending the proposed wavelet-based SR algorithms to incorporate with learning-based SR algorithms in a comprehensive method for further improving the efficiency of SR. Learning-based methods have attracted more and more interests due to their fine performance, but they require significant training data set. This future work will advance SR technology by improving the reconstruction performance meanwhile reducing the dependency on large data set.
- Improving the developed SR algorithms to handle the selection of threshold values and pre-set parameters in a smarter way.
- Developing a new wavelet-based video SR method for increasing the spatial resolution when the camera is moving and the observed object is moving together. A potential advantage of the proposed method is to have wider applications in video surveillance, such as security monitoring by increasing the spatial resolution meanwhile tackling the scene changes caused by weather conditions.

REFERENCES

[1] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview,” *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[2] J. Tian and K.-K. Ma, “A survey on super-resolution imaging,” *Signal, Image Video Process.*, vol. 5, no. 3, pp. 329–342, Sep. 2011.

[3] H. Zhang, Z. Yang, L. Zhang, et al., “Super-resolution reconstruction for multi-angle remote sensing images considering resolution diferences,” *Remote Sens.*, vol. 6, pp. 637–657, 2014.

[4] R. M. Willett, I. Jermyn, R. D. Nowak, et al., “Wavelet-based super-resolution in astronomy,” *Astronomical Data Analysis Software & Systems XIII*, vol. 314, pp. 107, 2004.

[5] F. S. Robinson, M.D; Chiu, S.J., Lo J.Y., Toth C.A., Izatt J.A, “Novel applications of super-resolution in medical imaging,” in *Super-Resolution Imaging*, P. Milanfar, Ed. CRC Press, 2010, pp. 383–412.

[6] J. Kamenicky, M. Bartos, J. Flusser, B. Mahdian, J. Kotera, A. Novozamsky et al., “PIZZARO: Forensic analysis and restoration of image and video data,” *Forensic Sci. Int.*, vol. 264, pp. 153–166, 2016.

[7] L. Zhang, H. Zhang, H. Shen, and P. Li, “A super-resolution reconstruction algorithm for surveillance images,” *Signal Processing*, vol. 90, no. 3, pp. 848–859, Mar. 2010.

[8] L. Yue, H. Shen, J. Li, Q. Yuan et al., “Image super-resolution: The techniques, applications, and future,” *Signal Processing*, vol. 128, pp.389–408, 2016.

[9] K. Nasrollahi and T. B. Moeslund, “Super-resolution: a comprehensive survey,” *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1423–1468, 2014.

[10] A. J. Shah and S. B. Gupta, “Image super-resolution-a survey,” in *2012 1st International Conference on Emerging Technology Trends in Electronics, Communication & Networking, Gujrat, 19-21 December,* 2012, pp. 1–6.

[11] J. Yang and T. Huang, “Image Super-Resolution: historical overview and future challenges,” in *Super-resolution imaging*, P. Milanfor, Ed. 2010.

[12] S. Chaudhuri and M. V. Joshi, *Motion-Free Super-Resolution*. Berlin: Springer, 2005.

[13] B. Zitová and J. Flusser, “Image registration methods: A Survey,” *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.

[14] P. Vandewalle, S. Süsstrunk, and M. Vetterll, “A frequency domain approach to registration of aliased images with application to super-resolution,” *EURASIP J. Appl. Signal Processing*, vol. 2006, pp. 1–14, 2006.

[15] L. Baboulaz and P. L. Dragotti, "Exact feature extraction using finite rate of innovation principles with an application to image super-resolution," *IEEE Transactions on Image Processing,* vol. 18, no. 2, p. 281–298, 2009.

[16] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov," *IEEE Transactions on Signal Processing,* vol. 46, no. 4, p. 886–902, April 1998.

[17] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial Applied Mathematics, Philadelphia, Pa, 1992.

[18] N. Nguyen and P. Milanfar, “A Wavelet-based interpolation-restoration method for superresolution (wavelet superresolution),” *Circuits Syst. Signal Process.*, vol. 19, no. 4, pp. 321–338, 2000.

[19] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Englewood Cliffs: NJ: Prentice Hall, 2007.

[20] S. Borman and R. L. Stevenson, “Spatial resolution enhancement of low-resolution image sequences: A comprehensive review with directions for future

research," Lab. Image and Signal Analysis., University of Notre Dame, Tech. Rep., 1998.

[21] S. Borman and R. L. Stevenson, "Super-resolution from image sequences- A review," *in Proce. 1998 Midwest Symp.Circcuits and Systems*, pp. 374–378, 1999.

[22] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Comuter Vision And Image Processing*, vol. 1, London: JAI Press, 1984, pp. 317–339.

[23] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust.*, vol. 38, no. 6, pp. 1013–1027, 1990.

[24] S. P. Kim and W. Y. Su, "Recursive high resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Processing.*, vol. 2, pp. 534–539, Oct.1993.

[25] H. Ur and D. Gross, "Improved resolution from subpixel shifted pictures," *CVGIP Graph. Model. Image Process.*, vol. 54, no. 2, pp. 181–186, 1992.

[26] A. Papoulis, "Generalized sampling theorem," *IEEE Trans. Circuits Syst.*, vol. 24, no. 6, pp. 652–654, Nov. 1977.

[27] J. I. Brown, "Multi-channel sampling of low pass signals," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 101–106, Feb. 1981.

[28] M. S. Alam, J. G. Bognar, R. C. Hardie and B. J. Yasuda, "Infrared image registration and high resolutionreconstruction using multiple shifted aliased video frames," *IEEE Trans. Instrum. Mcas.*, vol. 49, pp. 915–923, Oct. 2000.

[29] T. Q. Pham, L. J. Van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP J. Appl. Signal Processing*, vol. 2006, pp. 1–12, 2006.

[30] H. Knutsson and C. F. Westin, "Normalized and differential convolution," *in Processing of IEEE Computer Society Confernce on Computer Vision and Pattern Recoginsion (CVPR 93)*, pp. 515–523, New York, NY, USA, June 1993.

[31] S. Peleg, D. Keren, and L. Schweitzer, "Improving image resolution using subpixel motion," *Pattern Recognit. Lett.*, vol. 5, no. 3, pp. 223–226, 1987.

[32] M. Irani and S. Peleg, "Super resolution from image sequences," in *[1990] Proceedings. 10th International Conference on Pattern Recognition*, 1990, vol. ii, pp. 115–120.

[33] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP Graph. Model. Image Process.*, vol. 53, no. 3, pp. 231–239, 1991.

[34] M. Irani and S. Peleg, "Image sequence enhancement using multiple motions analysis," in *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992, pp. 216–221.

[35] M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, no. 4, pp. 324–335, 1993.

[36] P. L. Combettes, "The foundations of set theoretic estimation," in *Proceedings of The IEEE*, 1993, vol. 81, no. 2, pp. 182–208.

[37] H. Stark and P. Oskoui, "High resolution image recovry from image planearrays, using convex projections," *J. Opt. Soc. Am. A.*, vol. 6, pp. 1715–1726, 1989.

[38] A. M. Tekalp, M. K. Eren, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration," *in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), San Francisco, CA.*, vol. 3, no. 10, pp. 169–172, Mar.1992.

[39] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1064–1076, 1997.

[40] P. E. Eren, M. I. Sezan, and A. M. Tekalp, “Robust, object-based high-resolution image reconstruction from low-resolution video,” *IEEE Trans. Image Process.*, vol. 6, no. 10, pp. 1446–1451, 1997.

[41] A. J. Patti and Y. Altunbasak, “Artifact Reduction for Set Theoretic Super Resolution Image Reconstruction with Edge Adaptive Constrraints and Higher-Order Interpolants,” *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 179–186, 2001.

[42] B. C. Tom and A. K. Katsaggelos, “Reconstruction of a high-resolution image by simulation low-resolution images,” in *Proceedings, International Conference Image Processing*, 1995, no. 6, pp. 539–542.

[43] D. Capel and A. Zisserman, “Automated Mosaicing with Super-Resolution Zoom,” in *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, 23-25 June 1998*, 1998, pp. 885–891.

[44] R. R. Schulz and R. L. Stevenson, “A bayesian approach to image eExpansin for improved definition,” *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, 1994.

[45] R. R. Schultz and R. L. Stevenson, “Extraction of high-resolution frames from video sequences.,” *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, 1996.

[46] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, “Joint MAP registration and high-resolution image estimation using a sequence of undersampled images.,” *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1621–1633, 1997.

[47] H. Shen, L. Zhang, B. Huang, , and P. Li, “A Map approach for joint motion estimation, segmentation, and super resolution,” *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 479–490, 2007.

[48] X. Li,. X. Gao, Y. Hu., D. Tao, and B. Ning, “A multi-frame image super-resolution method,” *Signal Processing*, vol. 90, no. 2, pp. 405–414, 2010.

[49] D. Capel and A. Zisserman, “Super-resolution enhancement of text image sequences,” in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, 3-7 September 2000, Barcelona, Spain*, 2000, vol. 1, pp. 600–605.

[50] M. Elad and A. Feuer, “Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images,” *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, 1997.

[51] N. Nguyen, P. Milanfar, and G. Golub, “A Computationally efficient superresolution image reconstruction algorithm,” *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 573–583, 2001.

[52] M. Elad and Y. Hel-or, “A fast super-resolution reconstruction algorithm for pure translational motion and common space-invaiant blur,” *IEEE Trans. Image Process.*, vol. 10, no. 8, pp. 1187–1193, 2001.

[53] A. Zomet, A. Rav-Acha, and S. Peleg, “Robust super-resolution,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, vol. 1, pp. I–645–I–650.

[54] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, “Fast and robust multiframe super resolution,” *IEEE Trans. IMAGE Process.*, vol. 13, no. 10, pp. 1327–1344, 2004.

[55] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, “Robust shift and add approach to super-resolution,” in *Proceedings of the 2003 SPIE Conference on Applications of Digital Signal and Image Processing*, 2003, vol. 5203, no. 3, pp. 121–131.

[56] S. G. Chang, Z. Cvetkovic, and V. M., “Resolution enhancement of images using wavelet transform extrema extrapolation,” *Proc. ICASSP ,95*, vol. 4, pp. 2379–2382, 1995.

[57] W. K. Carey, D. B. Chuang, and S. S. Hemami, “Regularity-preserving image interpolation,” *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1293–1297, 1999.

[58] K. Kinebuchi, D. D. Muresan, and T. W. Parks, “Image interpolation using wavelet based hidden Markov trees,” in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 3, pp. 1957–1960.

[59] Shubin Zhao, Hua Han, and Silong Peng, “Wavelet-domain HMT-based image super-resolution,” in *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, vol. 3, pp. II–953–6.

[60] A. Temizel, “Image resolution enhancement using wavelet domain hidden markov tree and coefficient sign estimation,” in *2007 IEEE International Conference on Image Processing*, 2007, pp. V – 381–V – 384.

[61] A. Temizel and T. Vlachos, “Wavelet domain image resolution enhancement,” *IEEE Proc.-Vis. Image Signal Process.*, 2006, pp. 25–30.

[62] A. Temizel and T. Vlachos, “Wavelet domain image resolution enhancement using cycle-spinning,” *Electron. Lett.*, vol. 41, no. 3, p. 119, 2005.

[63] T. Vlachos, “Image resolution upscaling in the wavelet domain using directional cycle spinning,” *J. Electron. Imaging*, vol. 14, no. 4, p. 040501, Oct. 2005.

[64] A. Temizel and T. Vlachos, “Wavelet domain image resolution enhancement using cycle spinning and edge modelling,” in *Signal Processing Conference, 2005 13th European*, 2005, pp. 1–4.

[65] T. H. Reeves and N. G. Kingsbury, “Prediction of coefficients from coarse to fine scales in the complex wavelet transform,” in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, vol. 1, pp. 508–511.

[66] H. Demirel and G. Anbarjafari, “Satellite image resolution enhancement Using Complex Wavelet Transform,” *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 123–126, Jan. 2010.

[67] M. Z. Iqbal, A. Ghafoor, and A. M. Siddiqui, “Satellite image resolution enhancement using dual-tree complex wavelet transform and nonlocal means,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 451–455, May 2013.

[68] P. Jagadeesh and J. Pragatheeswaran “Image resolution enhancement based on edge directed interpolation using dual tree complex wavelet transform”, IEEE International Conference on Recent Trends in Information Technology, ICRTIT June 2011.

[69] P.-S. Tsai and T. Acharya, “Image up-sampling using discrete wavelet transform,” in *Proceedings of the 9th Joint Conference on Information Sciences (JCIS)*, 2006.

[70] G. Anbarjafari and H. Demirel, “Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image,” *ETRI J.*, vol. 32, no. 3, pp. 390–394, Jun. 2010.

[71] H. Demirel and G. Anbarjafari, “Discrete wavelet transform-based satellite image resolution enhancement,” *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 1997–2004, Jun. 2011.

[72] H. Demirel and G. Anbarjafari, “Image resolution enhancement by using discrete and stationary wavelet decomposition,” *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1458–1460, May 2011.

[73] A. Yavariabdi, C. Samir, and A. Bartoli, “3D medical image enhancement based on wavelet transforms,” in *Conference: Medical Image Understanding and Analysis*, 2011, pp. 172–176.

[74] Izadpanahi, S. and Demirel, H., 2012, July. Multi-frame super resolution using edge directed interpolation and complex wavelet transform. In Image Processing (IPR 2012), IET Conference on (pp. 1-5). IET.

[75] S. Izadpanahi and C. Ozcinar, “DWT based resolution enhancement of video sequences,” no. 2000, 2013.

[76] G. Anbarjafari, S. Izadpanahi, and H. Demirel, “Video resolution enhancement by using discrete and stationary wavelet transforms with illumination compensation,” *Signal, Image Video Process.*, vol. 9, no. 1, pp. 87–92, Jan. 2015.

[77] S. Izadpanahi and H. Demirel, “Motion based video super resolution using edge directed interpolation and complex wavelet transform,” *Signal Processing*, vol. 93, no. 7, pp. 2076–2086, 2013.

[78] S. Izadpanahi and H. Demirel, “Motion block based video super resolution,” *Digit. Signal Process.*, vol. 23, no. 5, pp. 1451–1462, Sep. 2013.

[79] Hsieh Hou and H. Andrews, “Cubic splines for image interpolation and digital filtering,” *IEEE Trans. Acoust.*, vol. 26, no. 6, pp. 508–517, Dec. 1978.

[80] R. Keys, “Cubic convolution interpolation for digital image processing,” *IEEE Trans. Acoust.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.

[81] J. A. Parker, R. V. Kenyon, and D. E. Troxel, “Comparison of Interpolating Methods for Image Resampling,” *IEEE Trans. Med. Imaging*, vol. 2, no. 1, pp. 31–39, Mar. 1983.

[82] E. Maeland, “On the comparison of interpolation methods,” *IEEE Trans. Med. Imaging*, vol. 7, no. 3, pp. 213–217, 1988.

[83] T. M. Lehmann, C. Gönner, and K. Spitzer, “Survey: Interpolation Methods in Medical Image Processing.,” *IEEE Trans. Med. Imaging*, vol. 18, no. 11, pp. 1049–1075, 1999.

[84] A. S. Glassner, K. Turkowski, and S. Gabriel, “Filters for common resampling tasks,” *Graph. Gems. New York Acadimic*, no. I, pp. 147–165, 1990.

[85] J. P. Allebach and P. W. Wong, “Magnifiying digital image using edge mapping,” U. S. Patent No.5,446,804, filed 14 April 1994, granted 29 August 1995, assignee: Hewlett-Packard Company.

[86] J. P. Allebach and P. W. Wong, “Edge-directed interpolation,” in *Proceedings of 3rd IEEE International Conference on Image Processing*, vol. 3, pp. 707–710.

[87] X. Li and M. T. Orchard, “New edge-directed interpolation,” *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, 2001.

[88] W.-S. Tam, C.-W. Kok and W.-C. Siu, "A modified edge directed interpolation for images", Journal of Electronic Imaging, Vol.19 (1), 013011, Jan-March 2010.

[89] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing,* vol. 5, no. 8, p. 1266–1271, 1996.

[90] B. Marcel, M. Briot and R. Murrieta, "Calcul de translation et rotation par la transformation de Fourier," *Traitement du Signal,* vol. 14, no. 2, pp. 135-149, 1997.

[91] L. Lucchese and G. M. Cortelazzo, "A noise-robust frequency domain technique for estimating planar roto-translations," *IEEE Transactions on Signal Processing,* vol. 48, no. 6, p. 1769–1786, 2000.

[92] W. K. Pratt, “Correlation techniques of image registration,”IEEE Transactions on Aerospace and Electronic Systems, vol. ASE-10, no. 3, pp.353–358, May. 1974.

[93] J. R. Bergan, P. Anandan, K. J. Hanna, and R. Hingorani, “Hierarchical model-based motion estimation,” in *Proceedings. of 2nd European Conference on Computer Vision*, Lecture Notes in Computer Science, pp. 237–252, Santa Margherita Ligure, May 1992.

[94] D. Keren, S. Peleg, and R. Brada, “Image Sequence Enhancement Using Sub-Pixel Displacements,” in *Proceedings CVPR ’88: The Computer Society Conference on Computer Vision and Pattern Recognition*, 1988, no. 3, pp. 742–746.

[95] M. Irani, B. Rousso and S. Peleg, “Computing occluding and transparent motions,” *International Journal of Computer Vision*, vol. 12, no. 1, pp. 5–16, 1994.

[96] J. Gluckman, “Gradient field distriution for the registration of images,” in *Proceedings of IEEE International Conference on Image Processing (ICIP 03)*, vol. 3, pp. 691–694, Barcelona, Spain, Sep 2003.

[97] D. Capel and A. Zisserman, “Computer vision applied to super-resolution ,” *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 75–86, 2003.

[98] M. A. Fischler and R.C.Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartogrphy,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[99] L. Y. Li, W. M. Huang, I. Y. H. Gu, and Q. Tian, “Statistical Modeling of Complex Background for Foreground Object Detection,” *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, 2004.

[100] W. Hu, T. Tan, L. Wang, and S. Maybank, “A survey on visual surveillance of object motion and behaviors,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 34, no. 3, pp. 334–352, 2004.

[101] S.-C. Huang, “An advanced motion detection algorithm with video quality analysis for video surveillance images,”IEEE Transactions on Circuits and Systems for video Technology, vol. 21, no. 1, pp.1–13, Jan. 2011.

[102] F.-C. Cheng, S.-C. Huang, and S.-J. Ruan “Scene analysis for object detection in advanced surveillance systems using laplacian distrubiution model,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 41, no. 5, pp.589–598, Sep. 2011.

[103] M. Oral and U. Deniz, “Centre of mass model - A novel approach to background modelling for segmentation of moving objects,” *Image Vis. Comput.*, vol. 25, pp. 1365–1376, 2007.

[104] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: real-time tracking of the human body,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, 1997.

[105] A. Manzanera and J. Richefeu, “A robust and computationally efficient motion detection algorithm based on sigma–delta background estimation,” *Proc. Fourth Indian Conf. Comput. Vision, Graph. Image Process.*, pp. 46–51, 2004.

[106] C. Liu and D. Sun, "On Bayesian adaptive video super resolution," *IEEE transactions on pattern analysis and machine intelligence,* vol. 36, no. 2, pp. 346-360, 2014.

[107] R. Liao, X. Tao, R. Li, Z. Ma and J. Jia, "Video Super-Resolution via Deep Draft-Ensemble Learning," *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 531-539, 2015.

[108] G. de Haan, “Progress in motion estimation for consumer video format conversion,” *IEEE Trans. Consum. Electron.*, 2000.

[109] E. Shechtman, Y. Capsi, and M. Irani, “Increasing space-time resolution in video,” *Proc. Eur. Conf. Comput. Vis.*, vol. 1, pp. 753–768, 2002.

[110] E. Shechtman, “Space-time super-resolution. Master’s Thesis,” *Master’s Thesis*, pp.1–44, 2003.

[111] E. Shechtman, Y. Caspi, and M. Irani, “Space-time super-resolution.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 531–545, 2005.

[112] O. Shahar, A. Faktor, and M. Irani, “Space-time super-resolution from a single video,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1–8, 2011.

[113] M. Shimano, . Okabe, I. Sato, and Y. Sato, “Video temporal super-resolution based on self-similarity,” *Adv. Comput. Vis. Pattern Recognit.*, pp. 411–430, 2013.

[114] M. Protter, M. Elad, H. Takeda, P. Milnfar, “Generalizing the non-local-means to super-resolution reconstruction,” *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 1958–1975, Jan. 2009.

[115] M.-H. Cheng, H.-Y. Chen, and J.-J. Leou, “Video super-resolution reconstruction using a mobile search strategy and adaptive patch size,” *Signal Processing*, vol. 91, no. 5, pp. 1284–1297, 2011.

[116] H. Takeda, P. Milanfar, M. Protter, and M. Elad, “Super-resolution without explicit subpixel motion estimation,” *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1958–1975, 2009.

[117] H. Takeda, S. Farsiu, and P. Milanfar, “Kernel regression for image processing and reconstruction. Master’s Thesis,” *Master’s Thesis*, vol. 16, pp. 349–66, 2007.

[118] R. Timofte, V. De Smet, L. Van Gool, "Anchored neighbourhood regression for fast example-based super-resolution," Proceedings, IEEE international Conference on Computer Vision (ICCV 2013), pp. 1920-1927, 2013.

[119] R. Timofte, V. De Smet, L. Van Gool, "A+: Adjusted anchored neighbourhood regression for fast super-resolution," Proceedings, IEEE Asian Conference on Computer Vision (ACCV 2014), 2014.

[120] J.-J. Huang and W.-c. Siu, "Learning hierarchical decision trees for single image super-resolution," IEEE Transaction on Circuits and Systems for Video Technology, 2015.

[121] C. Dong, C. C. Loy, K. He and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, No. 2, Feb, 2016.

[122] W. Yang, J. Feng, J. Yang, F. Zhao et al., " Deep edge guided recurrent residual learning for image super-resolution," Comput. Vision Pattern Recognit. ar Xiv:1701.05652, 2017.

[123] J. Yang, J. Wright, T. Huang and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing ,* vol. 19, no. 11, p. 2861–2873, 2010.

[124] L.-W. Kang, C.-C. Hsu, B. Zhuang, C.-W. Lin, and C.-H. Yeh, "Learning-based joint super-resolution and deblocking for a highly compressed image," *IEEE Trans. Multimed.*, vol. 17, no. 7, pp. 405–417, 2015.

[125] Y. Zhang, Y. Zhang, J. Zhang, and Q. Dai, "CCR: Clustering and collaborative representation for fast single image super-resolution," *IEEE Trans. Multimed.*, vol. 18, no. 3, pp. 405–417, 2016.

[126] J. Jiang, X. Ma, J. Zhang, C. Chen, T. Lu, Z. Wang and J. Ma, "Single image super-resolution via locally regularized anchored neighbourhood regression and nonlocal means," *IEEE Trans. Multimed.*, vol. 19, no. 1, pp. 15–25, 2017.

[127] A. K. Bhandari, V. Soni, A. Kumar, and G. K. Singh, "Cuckoo search algorithm based satellite image contrast and brightness enhancement using DWT–SVD," *ISA Trans.*, vol. 53, no. 4, pp. 1286–1296, Jul. 2014.

[128] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal.*, vol. 11, no. 7, pp. 674–220, July 1989.

[129] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[130] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 36–58, 2001.

[131] B. E. Usevitch, "A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000," *IEEE Signal Process. Mag.*, vol. 18, pp. 22–34, September 2001.

[132] A. Pande and J.Zambreno, *Embedded Multimedia Security Systems*, Springer: Verlag, 2013.

[133] R. C. Gonzalez and R. E. Woods, *Digital image processing using Matlab*. Englewood Cliffs: NJ: Prentice Hall, 2004.

[134] T. Celik and T. Tjahjadi, “Image resolution enhancement using dual-tree complex wavelet transform,” *IEEE Trans. Geosci. Remote Sens.*, vol. 7, no. 3, pp. 554–557, July. 2010.

[135] S. Azam, F. Tuz Zohra, and M. M. Islam, “A state-of-the-art review on wavelet based image resolution enhancement techniques: Performance Evaluation Criteria and Issues,” *Int. J. Image, Graph. Signal Process.*, vol. 6, no. 9, pp. 35–46, Aug. 2014.

[136] M. S. D. Lakshmi, “Robust satellite image resolution enhancement based on interpolation of stationary wavelet transform,” *Int. J. Sci. Eng. Res.*, vol. 4, no. 6, p. 1365, 2013.

[137] J. A. Richards and X. Jia, Remote sensing digital image analysis: an introduction, 5rd ed. Berlin: Springer, 2013.

[138] N. Jayant and P. Noll, *Digital Coding of Waveform*. Englewood Cliffs: NJ: Prentice Hall, 1984.

[139] S. M. De Jong and F. D. Van Der Meer, *Remote sensing image analysis: including the spatial domain*. Dordrecht: Springer, 2004.

[140] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[141] X.-P. Zhang, “Thresholding neural network for adaptive noise reduction,” *IEEE Trans. Neural Networks*, vol. 12, no. 3, pp. 567–584, May 2001.

[142] Satellite Imaging Corporation, “Satellite Image Gallery.” [Online]. Available: http://www.satimagingcorp.com/gallery/. [Accessed: 01-Apr-2015].

[143] Z. Wang, C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility structural similarity," IEEE Transactions on Image Processing, vol. 13, No. 4, April, 2004.

[144] Z. Xiang and P. Ramadge, “Edge-preservation image regularization based on morphological wavelets and dyadic trees,” *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1548–1560, April, 2012.

[145] N. Otsu, “A threshold selection method from gray-level,” *IEEE Trans. Syst. Man Cybern. Part B*, vol. 9, no. 1, pp. 62–66, 1979.

[146] V. Bannore, *Iterative-Interpolation Super-Resolution* Image Reconstruction. Berlin: Springer, 2009.

[147] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli et al., “On between-coefficient contrast masking of DCT basis functions,” (in German), *CD-ROM Proceedings. of the Third international Workshop on Video Processing and Quality Metrices for Consumer Electronics VPQM-07*, 25.-26. 2007.