

Motion block based Video Super Resolution

Sara Izadpanahi

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy
in
Electrical and Electronic Engineering

Eastern Mediterranean University
September 2013, Gazimağusa,
North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Elvan Yılmaz
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Doctor of Philosophy in Electrical and Electronic Engineering.

Prof. Dr. Aykut Hocann
Chair, Department of Electrical and Electronic Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Doctor of Philosophy in Electrical and Electronic Engineering.

Assoc. Prof. Dr. Hasan Demirel
Supervisor

Examining Committee

-
1. Prof. Dr. Enis Çetin _____
 2. Prof. Dr. Gözde Bozdağı Akar _____
 3. Assoc. Prof. Dr. Hüseyin Özkaramanlı _____
 4. Assoc. Prof. Dr. Hasan Demirel _____
 5. Assist. Prof. Dr. Önsen Toygar _____

ABSTRACT

A multi-frame super resolution process can be used for enhancing the resolution of video frames by employing the information of consecutive low-resolution frames taken from almost the same scene. Most of these super resolution algorithms are only suitable for global motion model. Nevertheless, if a local motion pattern such as movements of some objects happens between the low resolution frames a global motion model cannot provide efficient performance. Considering this problem, we propose a novel super resolution framework, where the moving and static regions in video frames are processed separately. Occlusion is another issue, which is not considered in most of the video super resolution processes. This problem occurs when a new object appears or an object disappears in the video frames. The proposed motion-block based super resolution method not only offers a local motion model but also deals with the occluded areas in a proper way.

This thesis presents a new video super resolution technique, based on the motion and static areas of the low resolution video frames. In order to separate the motion and static blocks, a block motion estimation method is employed between a reference and its neighbouring frames. Among the motion blocks, the occluded blocks are identified using an adaptive threshold applied on each block individually. Structure-adaptive normalized convolution (SANC) reconstruction method is used to generate the high resolution static and motion blocks where discrete wavelet transform (DWT) based interpolation is used to produce the high resolution occluded blocks. The static and motion blocks are combined into a high resolution frame. Finally, a sharpening process

is performed on the high resolution frame in order to generate the super resolved high resolution output frame. The experimental results show that the proposed technique generates significantly better qualitative visual results as well as higher quantitative PSNR and SSIM than the state of the art video super resolution algorithms.

Keywords: Super resolution, resolution enhancement, multi-frame super resolution, video super resolution, motion estimation, local motion patterns.

ÖZ

Çok çerçeveli süper çözünürlük işlemi video dizilerinin çözünürlüğünü hemen hemen aynı ayardaki düşük çözünürlüklü görüntülerden yararlanarak iyileştirmekte kullanılabilir. Çoğu süper çözünürlük algoritmaları sadece evrensel hareket modeli için uygundur. Yine de, eğer yerel hareket şablonunda bazı nesnelere düşük çözünürlüklü çerçevelerde sadece yerel hareketleri olursa evrensel hareket modeli yeterli bir performans ortaya çıkarmaz. Bu problemi dikkate alarak, görüntü dizilerindeki hareketli ve sabit bölgelerin ayrı ayrı işlendiği yeni bir çözünürlük çerçeve modelini önermekteyiz. Kapanma çoğu süper görüntü çözünürlük işlemlerinde dikkate alınmayan başka bir problemdir. Bu problem görüntü dizilerinde yeni bir nesne oluşunca yada yok olunca oluşur. Önerilen blok tabanlı süper çözünürlük yöntemi sadece yerel hareket modeliyle değil kapanan alanlarla da uyumlu bir şekilde çalışmaktadır.

Bu çalışmada düşük çözünürlüklü görüntü dizilerinin hareketli ve sabit alanlarından oluşan yeni bir süper görüntü çözünürlük tekniği sunulmaktadır. Hareketli ve sabit blokları ayırmak için, referans noktası ile komşu dizinler arasında hareketli blok kestirimi yöntemi kullanılmaktadır. Hareketli blokların arasındaki kapanmış bloklar, her blok için uygun eşik değerleri kullanılarak belirlenir. Yapısal-uyarlanırlı düzgeleme konvolüsyon (SANC) geriçatım yöntemi yüksek çözünürlükte hareketli ve sabit bloklar üretmek için kullanılırken, interpolasyon tabanlı ayırık dalgacık dönüşümü (DWT) yüksek çözünürlüklü kapalı blok üretmek için kullanılmaktadır. Sabit ve hareketli bloklar

birleřtirilerek yksek znrlkl ereveler oluřturulmaktadır. Son olarak da bir bileme iřlemi yksek znrlkl ereve zerinde sper yksek znrlkl ıktı erevesini yaratmak iin uygulanmaktadır. Deneysel sonular nerilen yntemin literatrde yer alan video sper znrlk algoritmalarına gre grsel grnt kalitesi ve sayısal gstergeler ,PSNR ve SSIM gibi metrikler, aracılıęı ile daha bařarılı olduęunu ortaya koymaktadır.

Anahtar Kelimeler: Sper znrlk, znrlk iyileřtirme, oklu-ereve sper znrlk, video sper znrlk, hareket kestirimi, yerel hareket rntleri.

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to Assoc. Prof. Dr. Hasan Demirel for expanding my knowledge in Image Processing during his fruitful course and his constructive advice and willingness to share his insight and wisdom.

Many thanks for the head of department of Electrical and Electronic Engineering, Prof. Dr. Aykut Hocanin for his guidance and support during the editing of this thesis. I am also deeply indebted to my instructors in the Department of Electrical and Electronics Engineering at Eastern Mediterranean University for their guidance, and support throughout my studies. I am grateful for suggestions, comments, kindness and contributions from my husband Şevki Kandulu and his precious family.

This work is dedicated to my son Eren Kandulu, father, Kavous Izadpanahi and my mother, Shahnaz Bahmanyar for the love and support they provided me through my entire life and in particular, I must acknowledge my beloved husband Şevki Kandulu that without his love, encouragement and editing assistance, I would not have finished this thesis.

TABLE OF CONTENTS

ÖZ.....	v
LIST OF FIGURES	x
LIST OF TABLES.....	xiv
CHAPTER 1	1
INTRODUCTION	1
1.1 History of Super Resolution.....	3
1.2 Problem Definition	5
1.3 Contributions of the Dissertation.....	6
1.4 Overview of the Thesis	7
CHAPTER 2	10
RESOLUTION ENHANCEMENT METHODS	10
2.1 Background.....	10
2.2 Interpolation methods.....	12
2.2.1 Nearest neighbor interpolation	12
2.2.2 Bilinear interpolation	13
2.2.1 Bicubic interpolation.....	14
2.2.2 New Edge Directed interpolation.....	15
2.3 Multi-frame super-resolution	19
2.3.1 Frequency-Domain Image Super resolution Methods	22
2.3.2 Spatial-domain Super-resolution methods	29
CHAPTER 3	33
LOCALIZED SUPER-RESOLUTION TECHNIQUES	33
3.1 Introduction.....	33
3.2 Motion-based Localized Super Resolution using frame differences and Discrete Wavelet Transform (MSR).....	37
3.2.1 Discrete Wavelet Transform (DWT).....	40
3.2.2 Detection of motion and static region	42
3.2.3 Super Resolution Process	44

3.3	Motion-Block-based Localized Super Resolution using Complex Wavelet Transforms (MBSR)	44
3.3.1	Dual Tree Complex Wavelet Transform (DT-CWT)	45
3.3.2	Motion detection.....	49
3.3.3	Extraction and insertion of the motion blocks	53
3.3.4	Super Resolution Process	56
3.4	Motion-block-based Localized Super Resolution using New Edge Directed Interpolation and Complex Wavelet Transforms (MBSR using NEDI).....	59
3.5	Simulation Results and Discussions.....	63
CHAPTER 4		78
THE PROPOSED MOTION BLOCK BASED VIDEO SUPER-RESOLUTION		78
4.1	Introduction.....	78
4.2	Proposed resolution enhancement method.....	80
4.2.1	Pre- processing module	81
4.2.2	Motion block processing module.....	84
4.2.2.1	Initial motion block reconstruction	84
4.2.2.2	Final motion block reconstruction	90
4.2.3	Post processing module	92
4.3	Experimental Results	95
CHAPTER 5		109
CONCLUSION and FUTURE WORK		109
5.1	Conclusions.....	109
5.2	Future Work.....	110
REFERENCES.....		111

LIST OF FIGURES

Figure 2.1: Observation model for video HR reconstruction [37].	11
Figure 2.2: Four LR shifted and rotated images of face (125x125).	12
Figure 2.3: SR using [54] and [55] (250x250).	12
Figure 2.4: Nearest neighbour interpolation for non-integer coordinates [44].	13
Figure 2.5: Bilinear interpolation for non-integer coordinates [44].	14
Figure 2.6: Bicubic interpolation [44].	15
Figure 2.7: Illustration of interpolation using EDI [47].	16
Figure 2.8: (P) input image (Q) the interpolated image P [49].	18
Figure 2.9 : The results of different interpolations (with enlargement factor of 4) of (a) LR image, using (b) Bilinear interpolation (c) bicubic interpolation (d) EDI (e) NEDI. [48]	19
Figure 2.10: Multi-frame super-resolution process [51].	21
Figure 2.11: 2D plane transformation	21
Figure 2.12: a) Reference image b) $ F1(u) $: Fourier transform of image a c) Rotated image 34 degree d) $ F2(u) $ Fourier transform of image c rotated 34 degree.[43]	24
Figure 2.13: Pham et al. image reconstruction scheme [55].	29
Figure 2.14: Iterative Back-Projection Approach [4]	31
Figure 3.1: (a) a sinusoidal wave, (b) a wavelet.	35
Figure 3.2: The block diagram of the MSR method presented in [30]	39
Figure 3.3: Single level analysis filter bank for DWT.	40
Figure 3.4: (a) test image, (b) single level DWT decomposition of the test image. .	41
Figure 3.5: A multilevel decomposition of an image using DWT.	42
Figure 3.6: Block diagram for a 3-level DT-CWT [72].	46
Figure 3.7: Impulse response of dual-tree complex wavelets at 4 levels and 6 directions. (a) Real part. (b) Magnitude.	47

Figure 3.8: (a) Sample image for transformation. (b) The magnitude of the transformation. (c) The real part of the transformation [73].....	48
Figure 3.9: Block diagram of the motion detection method.	53
Figure 3.10: Two neighboring motion blocks. Measures are in pixels.....	54
Figure 3.11: Block diagram of the motion block extraction.	55
Figure 3.12: The block diagram of the MBSR method presented in [31].....	59
Figure 3.13: Block diagram of the MBSR using NEDI technique [31].....	60
Figure 3.14: PSNR and SSIM result of resolution enhancement of “Mother & daughter” video sequence obtained from Vandewalle and SANC SR versus MBSR using NEDI technique for 200 consecutive frames.....	67
Figure 3.15: Result of different SR methods on ”Mother & daughter” video frames (PSNR, SSIM in parenthesis),(a) Reference HR frame. (b) Input LR frame. (c) SR using [54] [55] (d) MBSR using NEDI technique.	69
Figure 3.16: Result of different SR methods on “Container” video frames (PSNR, SSIM in parenthesis), (a) Reference HR frame. (b) Input LR frame. (c) SR using [54] [55] (d) MBSR using NEDI technique.....	70
Figure 3.17: Result of different SR methods on “Akiyo” video frames (PSNR, SSIM in parenthesis), (a) Reference HR frame. (b) Input LR frame. (c) SR using [54] [55] (d) MBSR using NEDI technique.....	71
Figure 3.18: PSNR result of resolution enhancement of “Foreman” video sequence obtained from Protter et al. SR versus MBSR using NEDI technique for various frames.....	73
Figure 3.19: (a) Reference HR image (12 th frame) of Foreman video sequence. (b) Protter et al. super resolution (c) MBSR using NEDI technique.....	74
Figure 4.1: Block diagram of the proposed super resolution technique.	81
Figure 4.2: Two motion vectors between three low resolution input frames, corresponding to a motion block at the reference frame.	82
Figure 4.3: Detailed block diagram of the pre-processing module.....	83

Figure 4. 4: Block diagram of the motion block processing consisting of initial and final motion block reconstruction.	85
Figure 4. 5: Occluded area in frame 90 and 91 of Foreman video sequence. (a) and (b) show occluded areas in frames 90 and 91 respectively. (c) Super resolved area using Keren et al.[3] registration and SANC reconstruction[55]. (d) Interpolated area using DWT based interpolation [72].	86
Figure 4. 6: Three consecutive frames I_{n-1} , I_n and I_{n+1} are aligned on a common motion block. The blocks are expanded by 2 pixels from sides in each frame.	89
Figure 4. 7: A part of the 88 th frame of News with four motion blocks: (a) reference blocks (b) reconstructed blocks without block expansion (c) reconstructed blocks with block expansion.	89
Figure 4. 8: The boundary (yellow) pixels of two neighbor blocks filtered using averaging.....	93
Figure 4. 9: Block diagram of the sharpening process.....	95
Figure 4. 10: Frame by frame PSNR results of resolution enhancement of “Ice” video sequence obtained from various resolution enhancement methods versus proposed technique.	99
Figure 4. 11: A part of 100 th frame of Foreman video sequence: (a) Reference HR image, (b) Input LR image, (c) Resolution-enhanced image by NEDI, (d) Marcel registration and SANC reconstruction, (e) DWT, (f) proposed technique.....	100
Figure 4. 12: A part of 50 th frame of Container video sequence: (a) Reference HR image, (b) Input LR image, (c) Resolution-enhanced image by NEDI, (d) Keren reg. and SANC rec., (e) DWT, (f) proposed technique.	101
Figure 4. 13: A part of 3 rd frame of News video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by proposed technique (d) NEDI (e) Vandewalle registration and SANC reconstruction with incorrect registration.	103
Figure 4. 14: A part of 3 rd frame of News video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by NEDI (d)	

Vandewalle registration and SANC reconstruction with correct registration (e) proposed technique.....	104
Figure 4. 15: A part of 25 th frame of foreman video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by MBSR using NEDI (d) proposed technique.	106
Figure 4. 16: A part of 5 th frame of foreman video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by MBSR using NEDI (d) proposed technique.	107
Figure 4. 17: A part of 13 th frame of foreman video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by MBSR using NEDI (d) proposed technique.	108

LIST OF TABLES

Table 3 1: Average PSNR (dB) and SSIM values obtained after applying different motion block sizes in the MBSR using NEDI technique.....	64
Table 3.2: The average PSNR (dB) and SSIM values of different resolution enhancement methods on the test video sequences.	66
Table 3.3: The average PSNR (dB) and SSIM values of different resolution enhancement methods on the test video sequences.	75
Table 4. 1: Average PSNR of the motion compensated blocks with different sizes in the various test sequences.	96
Table 4. 2: Average PSNR of the reconstructed blocks with and without assigning weight.....	97
Table 4. 3: The average PSNR results of the super resolved test videos using different resolution enhancement methods.....	98
Table 4. 4: The average PSNR and SSIM results of the super resolved test videos using MBSR using NEDI and Proposed SR methods.	105

LIST OF SYMBOLS/ ABBREVIATIONS

$d'_2(x)$	Translated version of image $l_1(x)$
$\Delta(k)$	Locus
ε_a	Error of angles
$g_2(x,y)$	Translated and rotated version of $g_1(x,y)$
$\hat{\theta}_a$	Rotation angle
$\hat{\theta}_b$	Refinement of $\hat{\theta}_a$
$I(x+p, y+q)$	Bilinear interpolated value
k_x	Zero crossing line
k_y	Zero crossing line
$L_2(k)$	Fourier transform of reference image
$ L_2(k) $	Power spectrum of $L_2(k)$
R	Orthonormal rotation matrix
$R(\theta)$	Rotation matrix
σ	Standard deviation
x'	Translated data
\bar{x}	Mean
$A \oplus B$	Dilation
AT	Alarm Trigger

BM	Background Modeling
CAT	Computer Aided Tomography
CCD	Charge-Coupled Device
CMOS	Complementary Metal-Oxide-Semiconductor
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
DT-CWT	Dual Tree Complex Wavelet Transform
EDI	Edge Directed Interpolation
EM	Expectation Maximization
FFT	Fast Fourier Transform
GST	Gradient Structure Tensor
HR	High Resolution
IBP	Iterative Back Projection
IDWT	Inverse Discrete Wavelet Transform
ISO	International Organization for Standardization
LMS	Least Mean Square
LR	Low Resolution
LW/PH	Line Widths per Picture Height
MAP	Maximum A Posteriori
ML	Maximum Likelihood
MMA	Modified Moving Average

MMSE	Minimum Mean Squared Error
MRF	Markov Random Field
MSE	Mean Square Error
NC	Normalized convolution
NEDI	New Edge-Directed Interpolation
OE	Object Extraction
PDF	Probability Density Function
POCS	Projection Onto Convex sets
PSF	Point Spread Function
PSNR	Peak Signal-to-Noise Ratio
QMF	Quadrature Mirror Filter
RLS	Recursive Least Square
SANC	Structure Adaptive Normalized Convolution
SD	Steepest Descent
SR	Super Resolution
SSIM	Structural SIMilarity
USB	Universal Serial Bus
WT	Wavelet Transform

CHAPTER 1

INTRODUCTION

In an imaging system, the imaging acquisition device that is usually a Charge-Coupled Device (CCD) or a Complementary Metal-Oxide-Semiconductor (CMOS) active-pixel sensor limits the image spatial resolution. Generally, to capture two-dimensional image signals, these sensors are set in a two dimensional array. The spatial resolution of the captured image is obtained by the number of sensor elements per unit area or in other word the sensor size. Obviously, a higher density of the sensors results in a higher spatial resolution achievable of the imaging system. In contrast, an imaging system with insufficient sensors produces low-resolution images with blocky effects, as a result of the aliasing from low spatial sampling frequency.

A basic solution to enhance the spatial resolution of an imaging system is to reduce the sensor size in order to increase the sensor density. Nevertheless, decreasing the sensor size decreases the total light incident on each sensor, which ends up with a problem called shot noise. In addition, increasing sensor density or corresponding image pixel density increases the hardware cost of a sensor. Thus, the spatial resolution of an image that can be captured is restricted by the hardware limitation on the size of the sensor.

As the image sensors limit the spatial resolution of the image, the optics also limits the image details or the high-frequency bands, due to lens aberration effects, lens blurs (related with the sensor Point Spread Function -PSF), aperture diffractions, and optical blurring due to motion. In most real applications such as surveillance cameras and cell

phone built-in cameras, it is not practical to make imaging chips and optical components with high resolution image capturing ability because of its cost. Other limitations of the resolution of a surveillance camera are the camera speed and hardware storage. Furthermore, it is difficult to use high resolution sensors in other applications such as satellite imagery due to physical restrictions.

An alternative solution is to apply signal processing to post-process the captured degraded image in order to trade off computational cost with the hardware cost. There is crucial information in low resolution images, hardly visible to the human eye. However, simply magnifying an image causes blurring or blocking effect. A straightforward method is using an interpolation technique, which only adds pixels to sharpen the image. However, these methods are not able to recreate the detail information in the low resolution image since there is no additional information provided. This means, the quality of an interpolated image is very much limited because the lost frequency components cannot be recovered. To overcome these problems an effective and economic technique is Super-Resolution (SR) reconstruction.

Super-Resolution techniques generate high-resolution (HR) images from several observed low-resolution (LR) images by combining the non-redundant information of multiple low-resolution frames. This combination produces a high resolution image by increasing the high-frequency components and eliminating the degradations caused by the imaging process of the low-resolution camera. The subpixel shifts between LR images provide the required non-redundant information.

These subpixel shifts can be the result of uncontrolled motions between the imaging device and scene such as handshakes during capturing image, or due to controlled motions like the satellite imaging system orbits the earth with predefined speed and

path. In a SR process estimating subpixel shifts or motion parameters between images is called registration, where projecting the low-resolution image onto the high-resolution lattice is referred as reconstruction [1].

1.1 History of Super Resolution

Tsai and Huang [2] were the first people who demonstrate a super resolution algorithm. Their method was implemented in the frequency-domain. Keren et al. described spatial-domain based methods for both registration and restoration parts of super resolution algorithm. In the registration step, a global translation and rotation was considered and the restoration step had two stages [13]. Akar et al. [93] proposed different resolution enhancement methods to get the high definition colour images. The methods were suggested to beat the colour artifacts on super resolution image and decrease the computational complexity in HSV domain applications. The other work [94] was focused on the definition, implementation and analysis on well-known techniques of super resolution in order to understand the improvements of the super resolution methods over single frame interpolation techniques.

Irani and Peleg [4] demonstrated another method for solving the restoration problem in super resolution process, which was using iterative back-projection technique. The suggestion in [5] is a set-theoretic method for super-resolution restoration problem. Defining convex sets is the main effect that improves the results. In order to generate an enhanced image Ng et al. [6] progressed a controlled, regularized, total least squares. Their idea was taking into account that except noise, there exist translational errors in around each pixel. To overcome the problem of transform-based preconditioned system of equations, Ng and Bose [7] indicated that the errors caused by the displacement effect the convergence rate in an iterative method. Developing a simple restoration system for colourful images is their other

contribution [8]. To deal with the Tikhonov-regularized super-resolution Nguyen et al. [9] illustrated circulant block pre-conditioners. By using this method, they accelerate the conjugate gradient descent algorithm. Schultz et al. [10] proposed A Maximum A Posteriori (MAP) estimator with Huber-Markov random field (MRF) prior. In the other methods, a MAP-MRF-based super-resolution algorithm, using blur is involved [11]. They applied defocus cue to restore the intensity of the scene with good quality and the depth field. Elad and Feuer [12] employed a mixture of MAP, Maximum likelihood (ML) and projection onto convex sets (POCS) methods to solve the problem of super resolution for degraded images.

In a reconstruction based super-resolution approach Lin et al. [13] obtained the quantitative limits. To determine the up-sampling limits they used conditioning analysis of the coefficient matrix.

Baker and Kanade [14] and Freeman et al. [15] proposed that greater super-resolution could be achieved by taking advantage of local regularities inherent in natural images. Local groups of pixels in natural images have much less variability than they would have in randomly generated images. Such regularities can be used to predict more accurately the interpolated pixels from the ones in the original image and thus generate visually plausible fine spatial details in the expanded image.

Since research in [2] shows that spatial domain super-resolution methods are computationally expensive procedures, it is acceptable to start with a "rough guess" and to achieve successfully finer estimates. For example, Elad and Feuer [12] used different approximations to the Kalman filter and examined their performance. In particular, recursive least squares (RLS), least mean squares (LMS), and steepest descent (SD) were considered.

Reddy et al. [16] demonstrated a Fourier domain based registration method to align images. These images were translated and rotated version of a reference image. Using a log-polar transform of the magnitude of the frequency spectra, image rotation and scale can be converted into horizontal and vertical shifts.

Lucchese and Cortelazzo [17] developed a registration process in the frequency domain. The estimation of relative motion parameters between the reference image and each of the other input images are based on the Fourier domain properties [15].

1.2 Problem Definition

A super resolution method with less error is essential for the success of many applications. Various SR algorithms have been introduced for enhancing the resolution of images. Most of these super resolution methods are only suitable for global motion model. However, for a local motion like movements of some objects between the low resolution frames a global motion model cannot offer effective performance. In light of this scope, a novel motion block based video super resolution method is proposed and studied.

Appearing or disappearing an object, which is called occlusion in the video frames, is another problem in the video super resolution processes. The occlusion problem, which is often ignored, is an important problem which should be taken into account for improved quality in SR processes.

The objectives of the proposed motion block based super resolution technique can be listed as follows:

- 1) High performance in terms of PSNR and SSIM in comparison to conventional and state-of-art techniques.
- 2) Achieving better visual quality.
- 3) Robustness of the technique on the area of occlusions.
- 4) Block based processing which would be laying the foundations of using the macro blocks, which are already utilized in state of the art advanced video coding standards.

1.3 Contributions of the Dissertation

A new super resolution technique for enhancing the resolution of the degraded video sequences has been introduced. The main contributions of this thesis can be summarized as follows:

- 1) Improving the resolution of low resolution videos by localizing the movements through consecutive frames and processing them individually during the super resolution process.
- 2) Introducing a new block based processing using information taken from an optical flow estimation method.
- 3) Recognizing the occluded areas using an adaptive threshold and dealing with them in a proper manner to improve the quality of the generated high resolution frame.
- 4) Refining the generated high resolution frame by using a de-blocking and a wavelet based de-blurring method.
- 5) Utilizing a block based processing approach, which is capable of using the macro blocks that are already used in state of the art advanced video coding standards.

The proposed technique benefits from state of the art super resolution methods for enhancing the resolution of the static and motion parts of the low resolution frames.

In this work, after dividing each frame into blocks, each block is labelled as static, motion or occluded block to be treated differently through the super resolution process. Employing an appropriate way of resolution enhancement method for each kind of frame blocks results in an output frame with higher resolution.

1.4 Overview of the Thesis

A comprehensive background on super-resolution techniques is given at the beginning of Chapter 2. In multi-frame super-resolution methods, it is possible to extract the details from each individual image and combine them to reconstruct a single high-resolution one. Second chapter presents an inclusive survey of the multi-frame super-resolution along with some of the necessary background material.

Multi-frame SR methods have been trying to solve two independent and sequential steps, registration and reconstruction, both of which have an extensive literature [2-4,16-19]. It is therefore impossible to understand and approach the topic without a strong background in these areas. Hence, this work attempts to present overviews of these fields before implementation of the approaches.

Recently wavelet based approaches are emerging in various applications. A good review of applications of wavelets to signal processing can be found in [20, 21]. One of these applications is the contribution of the wavelet transform in image resolution enhancement. The basic idea behind wavelet-based resolution enhancement techniques is to preserve the information at the high-frequency subbands of the low resolution image [22,29]. In chapter 3, after a comprehensive bibliographic study of

the existing wavelet-based resolution enhancement techniques, three methods are proposed and introduced. Various experiments are performed, in order to evaluate the performances of these proposed methods. The first method performs the image resolution enhancement using Discrete Wavelet Transform (DWT) [30], while the other algorithms use Dual Tree Complex Wavelet Transform (DT-CWT).

In all three methods after decomposing the image into different subbands various resolution enhancement approaches are used to increase the resolution of these subbands. At the end of each section, the results of these algorithms are presented in results and discussions. Additionally, chapter 3 focuses on the motion-based localized super resolution of video sequences. Various motion based SR algorithms using different wavelet transforms have been developed and presented in this part of thesis [30, 31 and 32]. These methods attempt to improve the resolution of low resolution videos by localizing the movements through consecutive frames. Among the presented motion localized SR techniques, [31] is outperformed by other higher-order techniques in terms of accuracy and visual appearance of the warped images. The proposed method produces noticeably sharper images with less blocking effect. Corresponding SSIM and PSNR of these methods in different video frames shows that this approach has the best performance in order to enhance the quality of the low resolution video frames between other approaches.

Considering the occlusion problem and using an optical flow estimation method [34] chapter 4 presents the last proposed method which is, a new video super resolution technique [33] including three modules of pre-processing module, motion block processing module and post processing module. These three modules and their effects on the super resolution process have been discussed in details in Chapter 4. The performance of this SR technique is evaluated using different benchmark video

sequences. The results in terms of PSNR and SSIM values and visual evaluation demonstrated the superiority of the technique presented among the alternative methods in the literature. Finally, chapter 5 concludes the thesis.

CHAPTER 2

RESOLUTION ENHANCEMENT METHODS

2.1 Background

Resolution is the ability to identify details in an image. In this framework, we are mostly concerned in spatial resolution. In digital imaging, the expression spatial resolution often refers to the pixel density in an image. However, the effect of a low-pass filter on the resolution of an image is more than increasing the pixel numbers by repeating each pixel. The International Organization for Standardization (ISO) measures the visual resolution of a digital camera using line widths per picture height (LW/PH) which is the highest frequency pattern of dark and light lines where each individual line can still be visually resolved [35].

High-resolution image/video is required in most of the electronic imaging applications, since it contains more details that can be critical for that application. An Image processing approach attempts to generate a high resolution (HR) Image from one or more low resolution (LR) versions of it.

The image/video observation model is employed to relate the desired referenced HR image/frame to all the observed LR images/frames. Usually, the image acquisition process involves warping, followed by blurring and down-sampling to generate LR images from the HR image. The detailed observation model for video HR reconstruction model is illustrated in Figure 2.1. Let the original HR image be a vector form by $h = [h_1, h_2, \dots, h_{L1R1 \times L2R2}]^T$, where $L1R1 \times L2R2$ is the size of the original HR image. If $L1$ and $L2$ indicate the down-sampling factors in the horizontal and vertical directions, respectively, then, each observed LR image has the size $R1 \times R2$. Therefore, the LR image can be represented as $y_k = [y_{k,1}, y_{k,2}, \dots, y_{k,R1 \times R2}]^T$. Here, $k = 1, 2, \dots, N$,

which N is the number of LR images. Assuming that each observed image is affected by additive noise, the observation model can be formulized as

$$y_k = DB_k W_k h + n_k \quad (1)$$

where W_k and B_k are warp and blur matrices, with the same size of $L_1 R_1 L_2 R_2 \times L_1 R_1 L_2 R_2$, respectively. D is a $R_1 R_2 \times L_1 R_1 L_2 R_2$ down-sampling matrix, and n_k stands for the $R_1 R_2 \times 1$ noise vector. Note that all the images have the same blurring function [36].

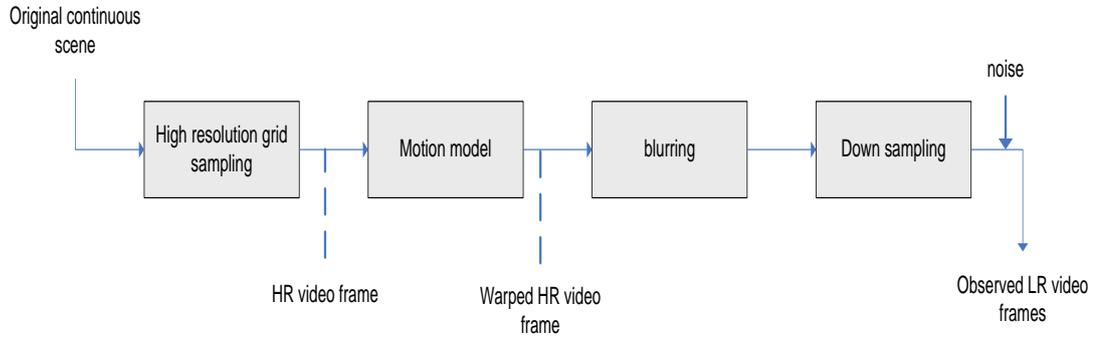


Figure 2. 1: Observation model for video HR reconstruction [37].

There are different approaches to increase the resolution of the blurred, downsampled images such as linear interpolator, adaptive image interpolation algorithms and edge directed interpolation techniques. However, a reasonable approach is to use signal processing techniques to obtain a high-resolution (HR) image from observed multiple low-resolution (LR) images. Such a resolution enhancement approach has been one of the most dynamic study areas, called multi frame super resolution (SR) [37-40]. Multi-frame super-resolution presents a way out to produce high-resolution images with finer details, by combining the information in a series of low-resolution frames, with relative sub-pixel shifts. It consists of two main phases:

first estimating motion parameters (registration) and then projecting the low resolution images onto a high resolution pattern (reconstruction).

Acquiring various images, at different times, from different point of views, and/or using different sensors, results in distorted images with respect to one another [2-4, 11, 12, 16-19, 41, 42]. The problem occurs when the information about this displacements are unknown. Image registration is the process of obtaining the best possible transformation matrix in which, it brings the distorted images back into spatial alignment. An accurate reconstruction of a high-resolution image is dedicated to a precise image registration [4, 16, 17, and 11]. Therefore, Image registration is the basis and also the challenging step of any multi-frame super resolution algorithm.

The image registration process is illustrated in Figure 2.2 and Figure 2.3. The first upper left image is the reference image and the three other images have been aligned with respect to the reference image.



Figure 2. 2: Four LR shifted and rotated images of face (125x125)

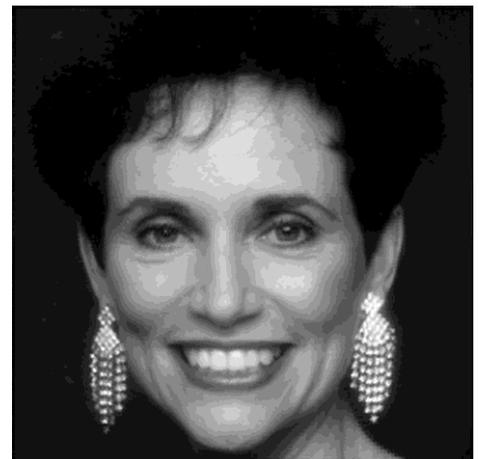


Figure 2. 3: SR using [54] and [55] (250x250)

This chapter targets an introduction to image resolution enhancement methods, by reviewing different interpolation and SR methods. After a discussion about Bicubic and edge directed interpolation methods, various SR algorithms consist of different registration and reconstruction methods will be explained.

2.2 Interpolation methods

The process of obtaining the values of a function at positions lying between its samples is called interpolation. This is achieved by fitting a continuous function through the discrete input samples. In this process, not only the input values defined at the sample points but also at arbitrary locations are evaluated.

In an image, interpolation determines the pixel values at non-integer coordinates by employing the pixel values at integer coordinates. The image quality highly depends on the applied interpolation technique. Various interpolation methods have been developed and can be found in the literature. Nearest neighbor, linear and bicubic interpolation are the most frequently used methods [3, 43].

2.1.1 Nearest neighbor interpolation

This method is also called as point pixel replication and shift algorithm. Nearest neighbor interpolation is the simplest method in comparison to the other interpolation methods. The pixel values of the interpolated image are determined using the value of their nearest sample point in the input image.

As shown in Figure 2.4, the projection of black point shown in image I to point $p1$ in image II can yield non-integer values. In this figure, four neighboring pixels around a pixel are shown. $p1$ is the pixel value of the nearest neighbor interpolated image I at pixel $p(2,1)$ since it is the nearest pixel to the black point in image I .

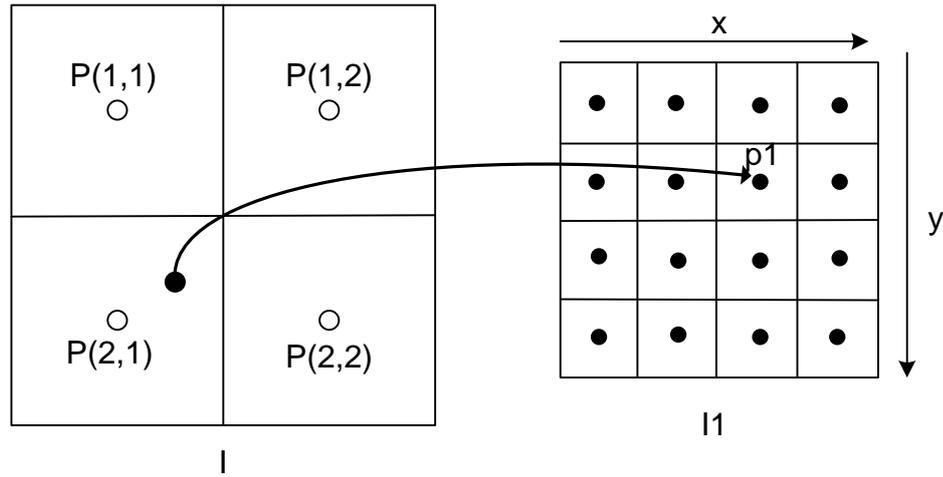


Figure 2. 4: Nearest neighbour interpolation for non-integer coordinates [44].

[45] Although nearest neighbor interpolation is the fastest and simplest methods and the simplest to implement, it has often the disadvantage of generating undesired artifacts such as, the distortion of stair-stepped effect around diagonal lines and curves and dropping or duplication of data values.

2.1.2 Bilinear interpolation

The output pixel value is assigned the value of a weighted average of pixels in the nearest 2-by-2 neighborhood in the input image. In figure 2.5, using non-integer values for x and y causes a mapping onto locations of the target grid. Therefore, it is necessary to involve the values at those locations based on the pixel values at integer coordinate locations. The bilinear interpolated value $pI(x, y)$ can be stated as:

$$pI(x, y) = (1 - dx)(1 - dy)p(1,1) + dy(1 - dx)p(2,1) + (1 - dy)dxp(1,2) + dx dy p(2,2) \quad (2)$$

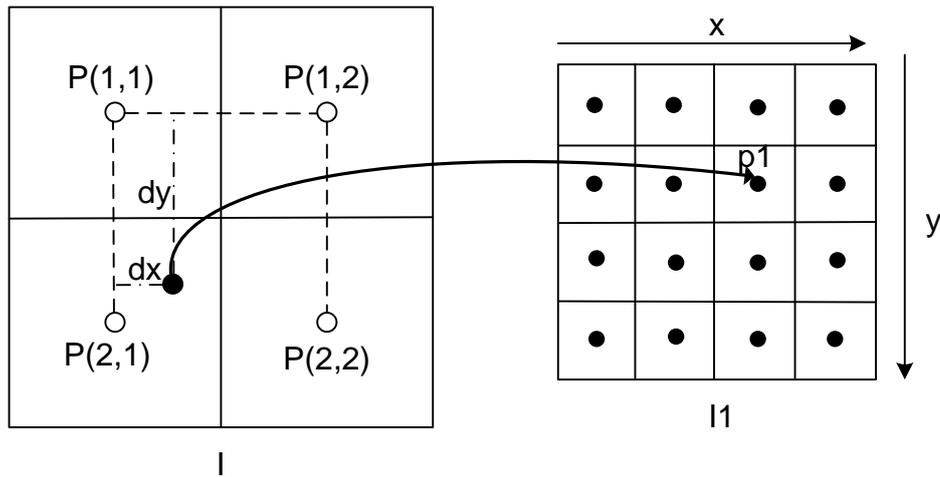


Figure 2. 5: Bilinear interpolation for non-integer coordinates [44].

The bilinear interpolated image is smoother than the nearest neighbor interpolated image.

2.2.1 Bicubic interpolation

As illustrated in Figure 2.6, the unknown pixel value is calculated from the value of a weighted average of the closest 4x4 neighbourhoods of known pixels for a total of 16 pixels. Given that, the known pixels are located at various distances from the unknown pixel, closer pixels are given a higher weighting than the further ones in producing Bicubic interpolated image.

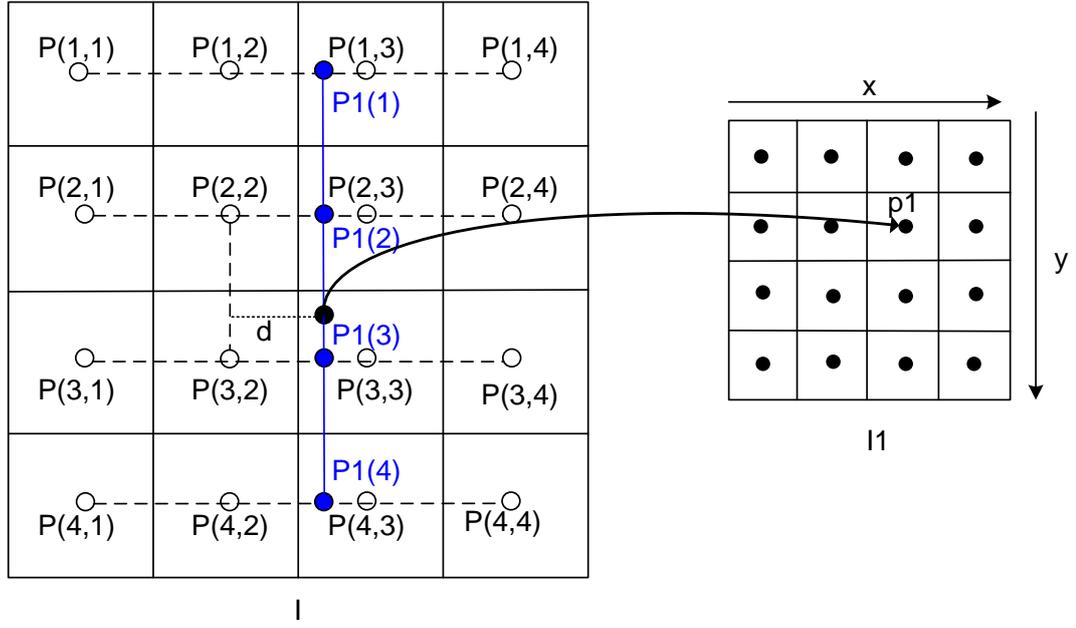


Figure 2. 6: Bicubic interpolation [44].

$$\begin{aligned}
 p1(k) = & p(k,1)(4 - 8(1+d) + 5(1+d)^2 - (1+d)^3) + p(k,2)(1 - 2d^2 + d^3) \\
 & + p(k,3)(1 - 2(1-d)^2 + (1-d)^3) + p(k,4)(4 - 8(2-d) + 5(2-d)^2 - (2-d)^3)
 \end{aligned} \quad (3)$$

The generated image is sharper compared to Bi-linear Interpolated one. However, it has less contrast in comparison to Nearest Neighbour interpolated image.

2.2.2 New Edge Directed interpolation

An interpolated image usually has problems in image edges, including the blurring of edges, blocking artifacts in diagonal directions and inability to generate fine details [46]. However, preserving edges is essential in many image applications. To solve these problems, Edge Directed Interpolation (EDI) is proposed by Allebach et.al [39].

Figure 2.7 shows an image, which contains three groups of pixels. This is a HR image, with the enlargement factor of 2, which is filled by EDI in three steps. First, dark pixels, which are the pixels in LR image, take place in the HR image from left to

right and up to down. In the next step, red pixels or pixels indexed by two odd values are determined as a weighted average of its four diagonal neighbors. Finally, white pixels, which are the rest of the pixels, are filled with its vertical and horizontal neighbors (red and dark pixels) by the same rule.

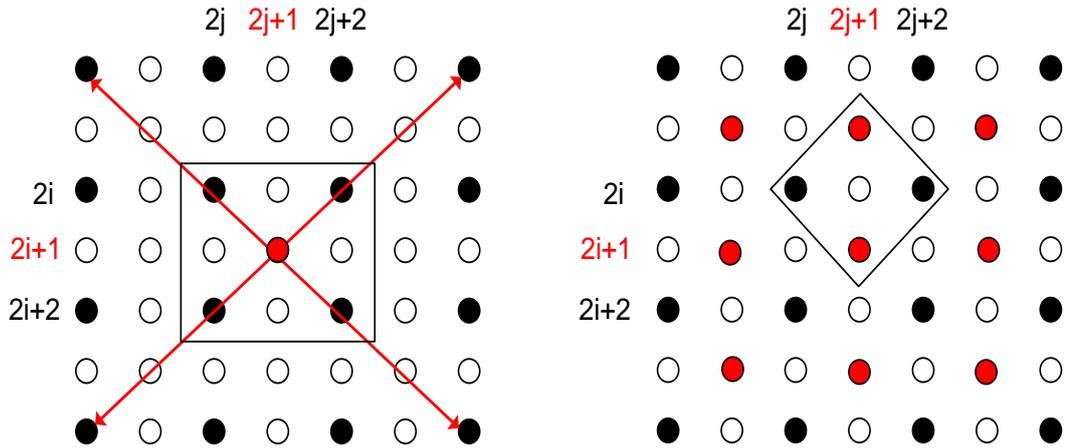


Figure 2. 7: Illustration of interpolation using EDI [47].

Li et.al [48] presented the idea of New Edge-Directed Interpolation (NEDI) which improved the performance of EDI. In NEDI method, no direction determination is considered, and the weights of new pixels are computed by assuming the local image covariance constant in a large window and at different scales. NEDI obtains a resolution enhanced image which is not smooth perpendicular to edges and is smooth parallel to edges.

Figure 2.8, illustrates image Q that is the interpolated version of input image P by the enlargement factor of 2. Each square in image P represents a single pixel. The last column and row of image Q is cropped away for easier explanation. As it is shown in Figure 2.8, image Q contains two types of pixels: 'a' and 'b'. Where 'a' pixels are located at $Q(2x, 2y) = P(x,y)$ and 'b' pixels correspond to those that are at $Q(2x+1,$

$2y+1$) in the interpolated image Q . NEDI consists of two steps. It first computes the values for 'b' pixels, and then for 'a' pixels. 'b' pixels are determined using their 4 known neighbor pixels. Afterwards, 'a' pixels are calculated using obtained 4 neighbor 'b' pixels. Assigning the calculated intensity values to 'a' and 'b' pixels results in a resolution enhanced image with sharp edges.

The low-resolution covariance can be easily estimated from a local window of the low-resolution image using the classical covariance method [48].

$$R = \frac{1}{M^2} C^T C, \quad r = \frac{1}{M^2} C^T yl \quad (4)$$

where $yl = [yl_1, yl_2, \dots, yl_{M^2}]^T$ is the data vector containing the $M \times M$ pixels inside the local window and C is a $4 \times M^2$ data matrix whose K th column vector is the four nearest neighbors of yl_k along the diagonal direction.

R is a 4×4 matrix, and r is a vector with 4 columns. According to Wiener filtering the optimal Minimum Mean Squared Error (MMSE) linear interpolation weights can be computed using the following.

$$\alpha = R^{-1}, \quad r = (C^T \cdot C)^{-1} (C^T \cdot y) \quad (5)$$

where, vector α , contains 4 interpolating weights for the 4 neighbor pixels. These weights are multiplied by the corresponding neighbor pixels and the results are added up to generate the new pixel.

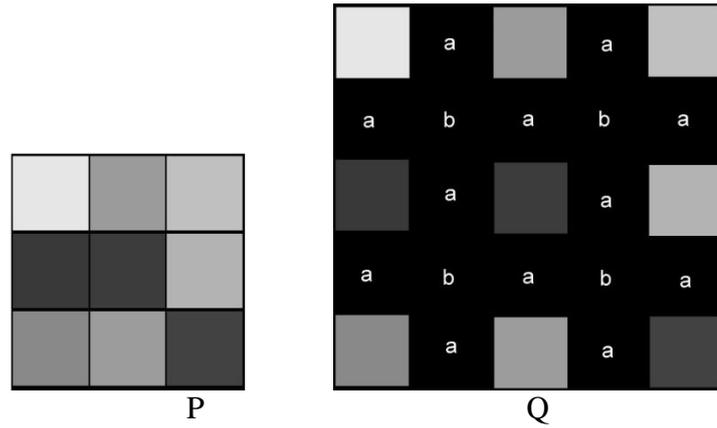


Figure 2. 8: (P) input image (Q) the interpolated image P [49].

Figure 2.9 shows a comparison between the results of an interpolated LR image using different interpolation methods. As you can see, new edge directed interpolated image has less blurred edges or artifacts around its edge area in comparison to Bicubic interpolated image. The generated artifacts in Bicubic interpolated image are the result of smoothing property of this method.

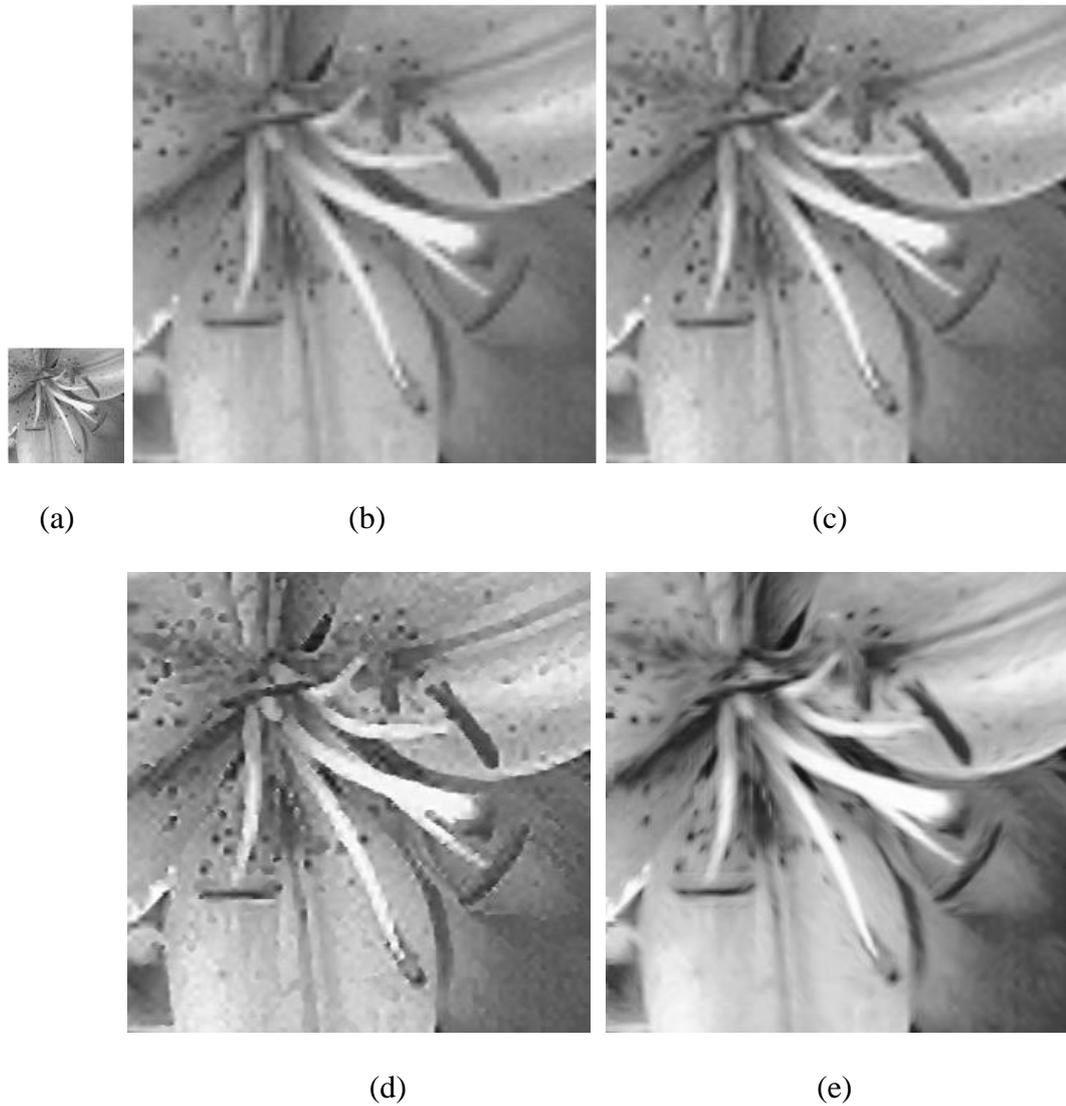


Figure 2. 9: The results of different interpolations (with enlargement factor of 4) of (a) LR image, using (b) Bilinear interpolation (c) bicubic interpolation (d) EDI (e) NEDI. [48]

2.3 Multi-frame super-resolution

Multi-frame super-resolution presents a way out to produce high-resolution images with finer details, by combining the information in a series of low-resolution images, with relative sub-pixel shifts. The information is embedded in the form of aliasing in this sequence of low-resolution images. The existing aliasing artifact in each

individual image degrades the visual quality of the image. A typical solution for this problem is applying a smoothing filter in the imaging sensor device prior to sampling. Smoothing filter causes blurring which is another degradation factor in images.

A less costly way to create a higher-quality image is utilizing the aliasing between the images. Usually, to reconstruct a single high-resolution image, these methods register several observed images to a common reference image in order to formulate multiple observed data. Thus, image registration process requires information of motion displacements involved in the observed image sequence. The unknown displacement information must be estimated from the observed image sequence to be employed in the reconstruction process [2, 16, 17, 50].

After registration another process is required in order to handle the resulting output grid with irregularly spaced sampling points. Therefore, any multi-frame super resolution is finalized by an image reconstruction process.

Figure 2.10 illustrates the graphical model of a multi-frame super-resolution process. A set of images are acquired from the same point of view with small movement using a single camera. The differences between these consecutive low-resolution images are used in the multi-frame super-resolution processes.

Different frequency and spatial-domain registration and reconstruction methods will be discussed in the next subsections.

Determining the proper mathematical model which relates the pixel coordinates in one image to pixel coordinates in another, is the difficulty of image alignment. In this part a basic 2D motion model and the transformation of a defined coordinates is reviewed.

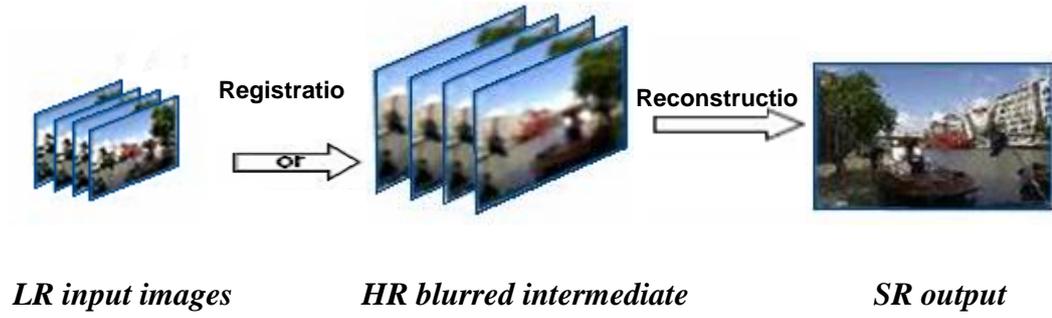


Figure 2. 10: Multi-frame super-resolution process [51].

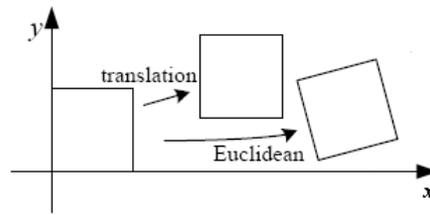


Figure 2. 11: 2D plane transformation

A 2D translations can be defined as $x' = x + t$ or

$$x' = h[It] \tilde{x} \quad (6)$$

where $\tilde{x} = (x, y)$ and I are the projective 2D coordinate and (2×2) identity matrix, respectively.

Another kind of transformation is considering rotation and translation in images. This transformation is also named as 2D rigid body motion or the 2D Euclidean transformation because of preserving Euclidean distances. This transformation is as follows:

$$x' = Rx + t \quad (7)$$

$$R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (8)$$

where R is an orthonormal rotation matrix with $RR^T = I$ and $|R|=1$.

2.3.1 Frequency-Domain Image Super resolution Methods

Typically, frequency-domain registration methods are based on the Fourier transform properties such as shifting and rotation. According to the shifting property of Fourier transform, rotation only changes the relationship between the amplitudes of two relatively warped versions of similar images. The amplitudes rotate with respect to each other at the origin of the spatial frequencies by the same angle as their spatial domain counterparts. Consequently, firstly the rotational component from the amplitudes of the Fourier transform is estimated and then, after compensating the rotation, and using phase correlation methods, the translational component is estimated [17].

Tsai and Huang [2] are the first researchers who present the analysis of Super resolution in frequency-domain. Their idea was extended by Kim et al. [41] by addressing the existing noise and blur during acquisition. The Expectation-Maximization algorithm demonstrated in [52], formulate an estimation of registration parameters.

Two frequency domain image registration methods [53, 54] are explained in this part. These algorithms are representative of a frequency domain method used in practice for image registration. The methods discussed utilize separability of rotational and translational components property of the Fourier transform [16] to model global

translational scene motion and take advantage of results from the sampling theory to affect super-resolution image registration from the data available in the observed image sequence.

2.3.1.1 Marcel et al. method

Marcel et al. [53] proposed an image registration approach which employed the Fourier domain properties to align images which are translated and rotated with respect to one another. They utilized the phase correlation methods to approximate camera movements under the assumption that these displacements are composed of translations and rotations in the imaging plane. The idea is to exploit the magnitudes of the Fourier transforms of the two images in polar coordinates. Consequently, two functions will be attained that differ in a translational displacement corresponding to the rotation angle. By applying a log-polar transform of the magnitude of the frequency spectra, image scale and rotation is converted into vertical and horizontal shifts which can be estimated using a phase correlation method. Their method uses the property of Fourier transform, which is the possibility of separating the rotational and translational components. According to this property, the translation only changes the phase data, while the rotation concerns with both phase and amplitude of the Fourier transform. A property of the 2D Fourier Transform is that, rotating the image causes the rotation of the spectrum in the same direction. Figure 2.12 shows an example of this property. In this example, figure 2.12 (c) is the rotated version of figure 2.12 (a) by 34 degrees. Accordingly, $|F_2(u)|$, which is the Fourier transform of figure 2.12 (c) is rotated in comparison to $|F_1(u)|$ (Fourier transform of figure 2.12 (a)) over the same angle as the spatial domain rotation.

Thus, by estimating and compensating the rotational component, using phase correlation techniques, the translational component is predicted.

Transforming $|F_1(u)|$ and $|F_2(u)|$ into polar coordinates, reduces the rotation over the angle α to a circular shift over α . Therefore, α can be calculated as the phase shift between $|F_1(u)|$ and $|F_2(u)|$. Hence, the image is transformed from Cartesian(x, y) grid into polar coordinates(r, α), for further rotation estimation.

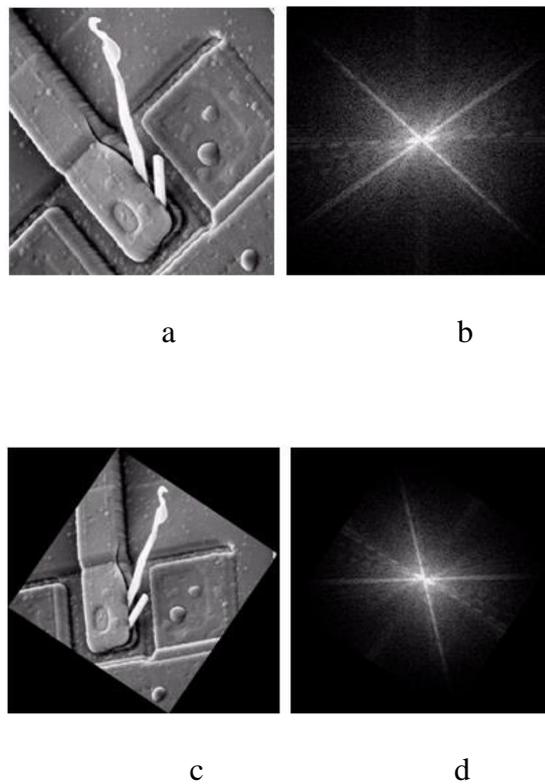


Figure 2. 12: a) Reference image b) $|F_1(u)|$: Fourier transform of image a c) Rotated image 34 degree d) $|F_2(u)|$ Fourier transform of image c rotated 34 degree.[43]

After estimating the rotation angle through phase correlation, they counter-rotate one of the two Fourier transforms and again apply a phase correlation technique to compensate the translation vector. Using this method the rotation angle is estimated

with an error less than the minimum discernible angle. The estimated error reported is about 0.9 degrees. In another experiment, they demonstrated that a minimum overlapping of 55% between the two images is necessary for their method to work.

Essentially for the low frequencies, which generally contain most of the energy, the interpolations are based on very few function values and thus introduce large approximation errors. An implementation of this method is also computationally intensive.

2.3.1.2 Vandewalle et al. method

Vandewalle et al.[54] presented a frequency domain method to estimate the motion parameters between a set of aliased images, based on their low-frequency, aliasing-free part. In their method only planar motion parallel to the image plane is used. The motion was defined as a function of three components: horizontal and vertical shifts, Δx_1 and Δx_2 , and a planar rotation angle ϕ .

Let $F_1(u)$ and $F_2(u)$ be the Fourier transforms of the reference signal $f_1(x)$ and shifted and rotated version of $f_2(x)$. They relate as

$$f_2(x) = f_1(R(x + \Delta x)) \quad (9)$$

$$R = \begin{bmatrix} \cos \Phi & -\sin \Phi \\ \sin \Phi & \cos \Phi \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \Delta x = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} \quad (10)$$

where, R is the rotation matrix.

Let $x' = x + \Delta x$, then the fourier domain expression of (4) is

$$F_2(u) = e^{-j2\pi u^T \Delta x} \int_x f_1(Rx') e^{-j2\pi u^T x'} dx' \quad (11)$$

The power spectrum or the amplitudes of the Fourier transforms are related as:

$$|F_2(u)| = |F_1(Ru)| \quad (12)$$

$|F_2(u)|$ is the rotated version of $|F_1(u)|$ over the same angle φ as the rotation between two images. It is well known that the power spectrums of two images do not depend on shift values Δx since the spatial domain translations affect only the phase values of the Fourier transform due to the shifting property of Fourier transform. Therefore, at first the rotation angle φ is calculated from the amplitudes of the Fourier transforms and then the translation Δx can be calculated by applying phase correlation methods.

The steps of Vandewalle et al. registration method is as follows:

Step 1: All low resolution images, $f_{LR,m}$ ($m=2, \dots, M$), are multiplied by a Tukey window for making them circularly symmetric. Where, M is the total number of low resolution images for registration. The resulting windowed images are called $f_{LR,w,m}$.

Step 2: $F_{LR,w,m}$ which is the Fourier transform of $f_{LR,w,m}$ is calculated.

Step 3: Rotation estimation: in this step the rotation angles between $f_{LR,w,m}$ and the reference image $f_{LR,w,1}$ are approximated as follows:

- (I) The polar coordinates, (r, θ) , of the $f_{LR,w,m}$ are calculated.
- (II) The average value $h_m(\alpha)$ of the fourier coefficient for every 0.1 degrees angle, α , is computed. Where, $\alpha-1 < \theta < \alpha+1$ and $0.1\rho < r < \rho_{max}$. Where, ρ is the image radius or half of the image size and ρ_{max} is set to 0.6.
- (III) The rotation angle, Φ_m , is estimated by finding the maximum of correlation between $h_m(\alpha)$ and $h_1(\alpha)$.
- (IV) The rotation of the image $f_{LR,w,m}$ is recovered by rotating it by $-\Phi_m$.

Step 4: Shift estimation: shift parameters consist of horizontal and vertical shifts between every images with respect to reference frame is approximated as below:

- (I) The phase difference between images in comparison to the reference image is computed as $\angle (f_{LR,w,m} / f_{LR,w,1})$.
- (II) The calculated phase differences with unknown slopes Δx , is used to define a plane for all frequencies $-u_s + u_{max} < u < u_s - u_{max}$
Where, u_s and u_{max} are sampled and maximum frequencies, respectively.
- (III) The shift parameters are estimated as the least square of the equations.

2.3.1.3 Structure Adaptive Normalized Convolution method

Pham et al.[55] presented a structure-adaptive algorithm based on the framework of normalized convolution (NC). This method applied for image fusion from irregularly sampled data. The local signal or in two dimensional image, is approximated through a projection onto a subspace spanned by set of basis functions. For improving signal-to-noise ratio and reducing diffusion across discontinuities, the window function of adaptive NC is adapted to local linear structures, so that more samples of the same modality gather for the analysis.

One of the methods for local signal modeling from projections onto a set of basis functions is Normalized convolution (NC) [56]. Generally, a polynomial basis $\{1, x, y, x^2, y^2, xy, \dots\}$ is used in this method. Where, the vectors, $1 = [1 \ 1 \ \dots \ 1]^T$ (N entries), $x = [x_1 \ x_2 \ \dots \ x_N]^T$, $x^2 = [x_1^2 \ x_2^2 \ \dots \ x_N^2]$, and etc. are constructed from local

coordinates of N input samples. Applying polynomial basis functions changes the traditional NC equivalent to a local Taylor series expansion. The intensity value at position $s = \{x + x_0, y + y_0\}$ within a local neighborhood centered at $s_0 = \{x_0, y_0\}$, is estimated by a polynomial expansion as follows:

$$\hat{f}(s, s_0) = p_0(s_0) + p_1(s_0)x + p_2(s_0)y + p_3(s_0)x^2 + p_4(s_0)xy + p_5(s_0)y^2 + \dots \quad (13)$$

where, $p(s_0) = [p_0 \ p_1 \ p_2 \ \dots \ p_m]^T(s_0)$ are the projection coefficient onto the corresponding polynomial basis functions at s_0 . $\{x, y\}$ are the local coordinates of sample s with respect to the center of analysis s_0 .

NC requires the signal certainty to be known. For this purpose, a Gaussian function forms the robust certainty as:

$$c(s, s_0) = \exp\left(-\frac{|f(s) - \hat{f}(s - s_0)|^2}{2\sigma_r^2}\right) \quad (14)$$

where, $f(s)$ and $\hat{f}(s, s_0)$ are measured and estimated intensities at position s , respectively.

In order to localize the polynomial fit, NC applies an applicability function, which is an isotropic, radially decaying function with a size proportioned to the scale of analysis. Pham et al. applied an adaptive anisotropic Gaussian kernel, which adapts its shape and orientation along the underlying image structure, for this purpose. To construct an adaptive kernel at an output pixel, an initial estimation of the output intensity is calculated. The initial estimation helps to approximate the gradient structure tensor (GST), [57] which contains the local image structure information such as ϕ and the anisotropy A . The adaptive applicability function is:

$$a_{2D}(s, s_0) = \rho_{2D}(s, s_0) \cdot \exp\left(-\left(\frac{x \cos \phi + y \sin \phi}{\sigma_u(s_0)}\right)^2 - \left(\frac{-x \sin \phi + y \cos \phi}{\sigma_v(s_0)}\right)^2\right) \quad (15)$$

where, ρ is a function centered at the origin and limits the kernel support to a certain radius and $s - s_0 = \{x, y\}$ are the local coordinates of input samples with respect to s_0 . σ_u and σ_v are the directional scales of the anisotropic Gaussian kernel.

Figure 2.13 depicts Pham et al. image reconstruction scheme. The registered LR images with their displacement parameters are fused in a fixed HR grid using robust and adaptive fusion method. A de-convolution is applied for de-blurring and de-noising the output image.

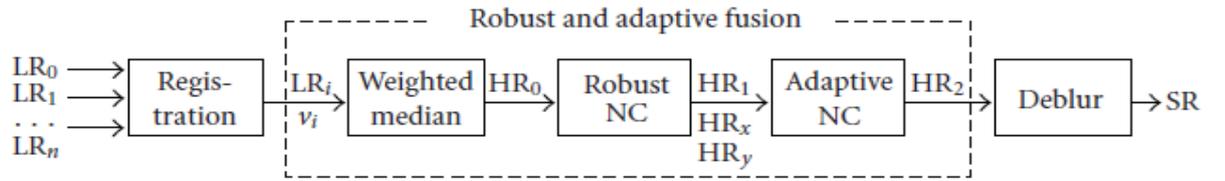


Figure 2. 13: Pham et al. image reconstruction scheme [55].

2.3.2 Spatial-domain Super-resolution methods

As mentioned another major category of multi-frame super-resolution registration methods are based on spatial domain formulation. The term spatial domain refers to the image plane itself, and the approaches manipulate the pixel of a given image for enhancement. In the other words, spatial domain methods are procedures that operate directly on pixels [43]. Their advantages include a great flexibility in the choice of motion model, motion blur, optical blur, and the sampling process. Another important factor is that the constraints are much easier to formulate.

The iterative methods are the most significant among the spatial domain methods, and are the focus of the present work. The most important advantages of an iterative technique lie in the ability to handle large image sequences, easy inclusion of a priori knowledge in the spatial domain, and the ability to handle spatially varying degradations.

There are many iterative methods to solve super-resolution reconstruction problems. Since previous researches [2] show that spatial domain super-resolution methods are computationally expensive procedures, it is acceptable to approach it by starting with a "rough guess" and achieving successfully finer estimates. For example, Elad and Feuer [12] use different approximations to the Kalman filter and examine their performance. In particular, recursive least squares (RLS), least mean squares (LMS), and steepest descent (SD) are considered.

As mentioned before, Keren, et al. [3] illustrated a spatial-domain based method to align the images by applying a global translation and rotation model. They proposed a two-step method restoration algorithm. As graphically illustrated in Figure 2.14 the image registration procedure assumes R observed images from which a gradient-based scheme estimates shift and rotation parameters for each of $R - 1$ images relative to a chosen reference image. With the image registration parameters estimated, the first stage of the restoration procedure registers the observed images and a high-resolution restoration grid is imposed on the "stack" of observed images. Each pixel in the high-resolution image is then chosen as the mean of the values in the set of observed image pixels whose centers fall within the area of the high-resolution pixel under consideration.

The Iterative Back projection (IBP) algorithm suggested by Irani and Peleg [4], which will be explained in detail in the following section, originated from computer-aided Tomography (CAT). The algorithm simulates the imaging process, back-projects the error between the simulated low-resolution images and the observed low-resolution images to the super-resolution image. Later in [58], they modify their method to handle more complicated motion types, which can include local motion, partial occlusion, and transparency. The fundamental back-projection scheme remains identical to the previous one, which is not very flexible in terms of incorporating a priori constraints on the solution area. Shah and Zakhor [59] use a reconstruction method similar to that of Irani and Peleg. They also propose a novel approach to motion estimation that considers a set of possible motion vectors for

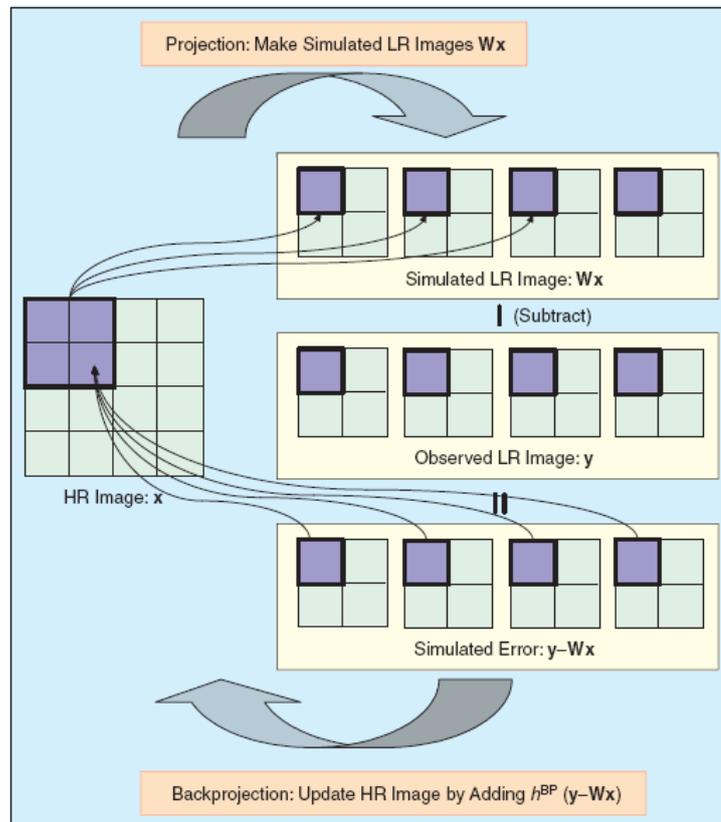


Figure 2. 14: Iterative Back-Projection Approach [4]

each pixel and eliminate those that are inconsistent with the surrounding pixels. In order to reduce the noise, Stark et al. applied a set theoretic algorithm, projection onto convex sets (POCS), to the super-resolution reconstruction [58]. It is convenient to integrate a priori information in POCS. However, the set theoretic algorithm suffers from non-uniqueness of the solution, slow convergence and high computational cost.

The main idea behind the registration is to detect accurate displacements between the sequences of images taken from the same point of view.

Irani et al. [4] proposed a spatial domain based motion estimation method, which uses translations and rotations. This method creates a group of low resolution images from original images; using Gaussian pyramids the set of input images are converted into multiple low pass filtered sets. Using the coarsest image sets, motion parameters are found and applied to finer sets. Thus it is an efficient and fast iterative back projection algorithm.

CHAPTER 3

LOCALIZED SUPER-RESOLUTION TECHNIQUES

3.1 Introduction

Determining motion parameters between consecutive frames is the goal of all multiframe registration tasks. Methods for performing these tasks are usually based on representing an image using global displacement in comparison to neighbour frames. The relevant representations and corresponding displacement parameter measures can vary significantly. Displacements are often represented using global shift and rotation estimators. However in many applications only shift parameter is sufficient for a qualified super resolution [17]. Although these measurements are variable, they share the same basic assumption that there exist a global shift and a global rotation difference between images. This assumption, however, may be so effective, we can clearly notice that there are some parts in image, which is not shifting or rotating in comparison to

other frames. In other words, there are stationary regions with no need to be shifted or rotated. In contrast, there exist regions in image with local displacements. What makes these regions distinct is the fact that each part of these portions is moving (shifting and rotating) in different direction. The effect of local motion is even stronger in the video sequences that the camera is stable and one or more objects are moving in the scene. Thereupon, separating motion regions from static part of the consecutive frames and registering only the motion part of images improves the ability to estimate the displacement parameters. The above process allows to do better registration. Nevertheless if each part of these objects moves in different orientation this process still suffers from the similar inherent problem. To work out this issue, using a motion detection algorithm, we detect the motion regions. Then the motion block extraction algorithm divides and extracts the motion parts into sufficient small blocks. This is mainly because we tend to have a single direction of motion in each block. These blocks can then be treated as if taken from different low resolution images for registration proposes. Therefore, separating motion regions from the static regions and processing them separately, results in a more accurate registration process. The motion regions go through multiframe super resolution for the localized resolution enhancement. On the other hand the static regions go through a resolution enhancement process.

The main loss of a video frame after being super resolved is on its high frequency components (i.e. edges), which is due to the smoothing caused within the super resolution processes. Therefore, in order to increase the quality of the super resolved video sequence, preserving the edges (high frequencies) of each frame can increase the

quality of the super resolved sequence. Recently, employing a wavelet transform in super resolution processes is a solution for this problem [60-62].

A small wave with its energy concentrated in time is called wavelet. It is an appropriate tool for transient, non-stationary or time-varying phenomena. One of the important properties of the wavelet is the ability to allow simultaneous time and frequency analysis [63, 64]. Figure 3.1 (a) shows a sinusoidal wave which is smooth predictable and everlasting. They are suitable for deterministic basis functions in Fourier analysis in order to expand a time invariant, or stationary signal function. A wavelet is illustrated in Figure 3.1 (b). A wavelet is of limited duration, irregular and sometimes asymmetric. They can be used as non-deterministic or deterministic basis to produce and analyze natural signals and achieve an accurate time-frequency representation. These types of analysis, which are not possible with waves using conventional Fourier analysis, are from the important characteristics of a wavelet.

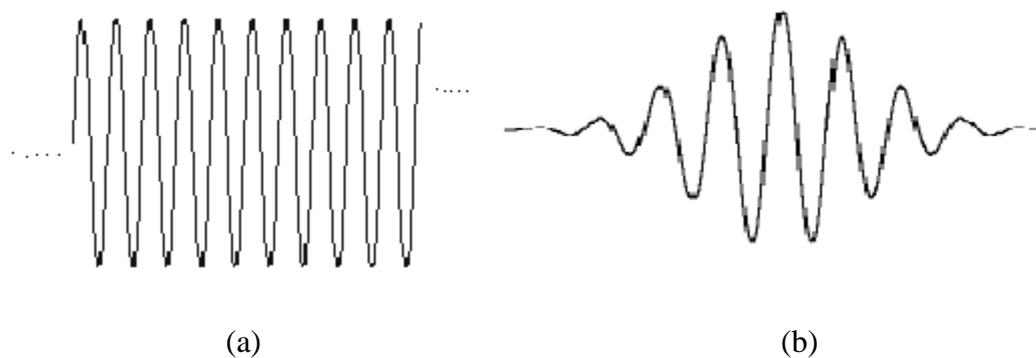


Figure 3. 1: (a) a sinusoidal wave, (b) a wavelet.

The wavelet analysis process adopts a “mother wavelet” or “analyzing wavelet” function prototype. A contracted, high frequency version of the prototype wavelet obtains the temporal analysis, where a dilated, low frequency version of the same wavelet achieves the frequency analysis [65]. An analogous to the Fourier transform

(FT) pair is the Wavelet Transform (WT) pair, which is obtained by a mathematical formulation of signal expansion using wavelet.

In [20,21] the authors explained the applications of wavelets to signal processing in details. An application of the wavelet transform is its contribution to image resolution enhancement. Wavelet-based resolution enhancement techniques improve the resolution of the given image by approximating the preserved information at its high-frequency subband [22-29]. Wavelet transform decomposes the image into different low and high frequency subbands. The key idea of these techniques is that, the intention is to estimate the high frequency subbands of wavelet transform and the resolution enhanced image is the low frequency subband amongst wavelet-transformed subbands of the original one. As a result, a resolution-enhanced image with more information at its high frequency subbands can be obtained.

Chang et al.[22] presented a wavelet based interpolation in which the coefficients of the high-frequency subbands were obtained using the regularity of edges across various scales. The method was successful in estimating the coefficients having significant magnitude whereas, approximating small coefficients were difficult. Later on, a Crouse et al.[26] demonstrated a hidden Markov model in order to predict the coefficients a high levels. The model represented the statistical relations among the coefficients at lower level. The problem occurred when the extracted information from a set of training images, was not efficiently matched with the input image. Another wavelet based resolution enhancement method was presented by Temizel et al. [28]. In this method, cycle spinning components were affected by the local edge direction. They recovered the problem of ringing artifacts caused by decimation, by averaging the translated zero-padded reconstructed images.

In this chapter, different localized super resolution techniques using various wavelet transforms are proposed and explained in details.

3.2 Motion-based Localized Super Resolution using frame differences and Discrete Wavelet Transform (MSR)

In this section, the method presented in [30] which applies a motion based localized super resolution, is demonstrated in details. In this thesis this method is abbreviated as MSR.

Super resolution is consist of two main parts, registration and reconstruction. As mentioned earlier registration is to estimate displacements (motion parameters) between two images. According to this property we can divide every image into two parts: motion regions and constant regions. For the region of frame which has no movement in comparison of last frames we can do only reconstruction because there is no displacement to estimate. This is the basic idea of MSR method which causes an appropriate registration. The illustrated method in [30] involves the following steps:

Step 1: determining motion regions and separating them from constant region.

Step 2: decomposing the static and motion regions into different frequency subbands using DWT.

Step 3: super resolving the subbands of motion regions and interpolating the subbands of constant region.

Step 4: applying IDWT to the subbands of motion and static region in order to generate super resolved static and motion regions.

Step 5: combining super resolved and interpolated region and produce the super resolved frame.

The block diagram of the algorithm of MSR method is shown in figure 3.2.

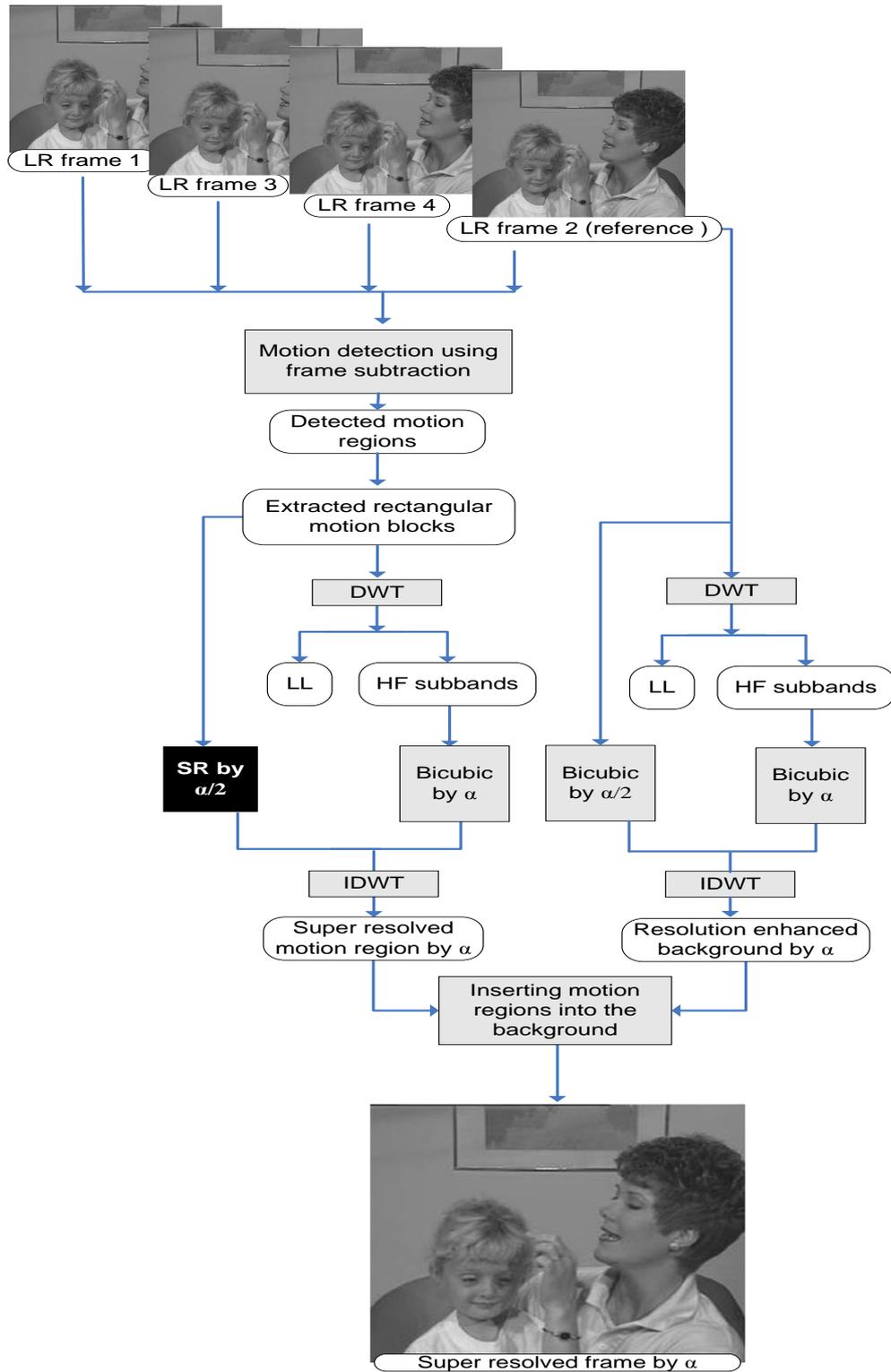


Figure 3. 2: The block diagram of the MSR method presented in [30]

3.2.1 Discrete Wavelet Transform (DWT)

The function $\psi(t)$ is called a mother wavelet. The family of a mother wavelet function is obtained by shifting and scaling the function as follows:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (16)$$

where, a is a real positive scaling factor and b is a real shifting factor

A filterbank is used to implement a single level DWT [66]. The detailed scheme of this implementation is shown in figure 3.3. As a result, three sub-images called HL, LH and HH, corresponding to horizontal, vertical and diagonal directions and a low resolution image called LL are produced. An example of this decomposition is shown in figure 3.4.

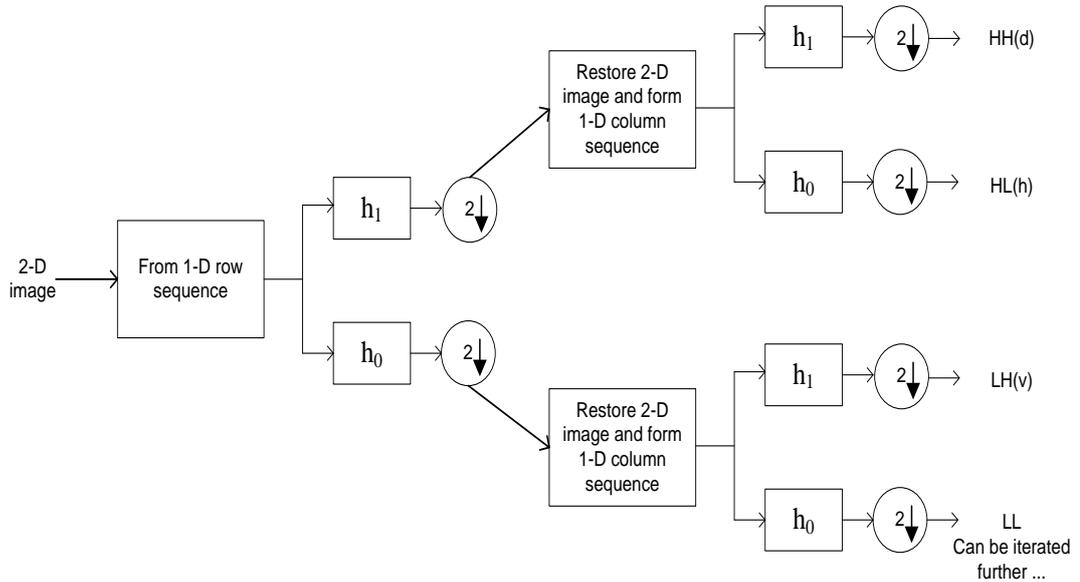
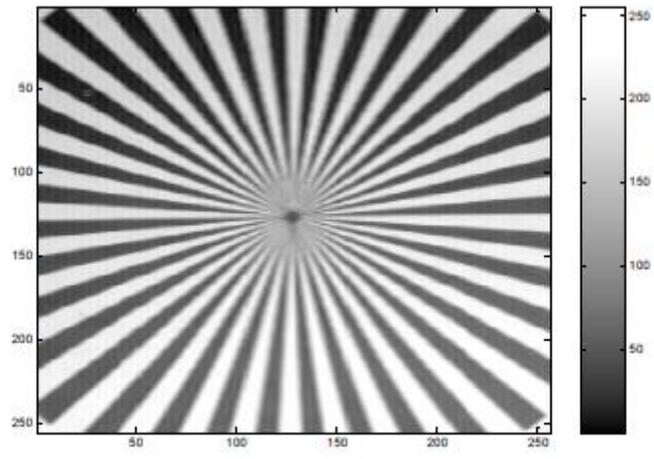
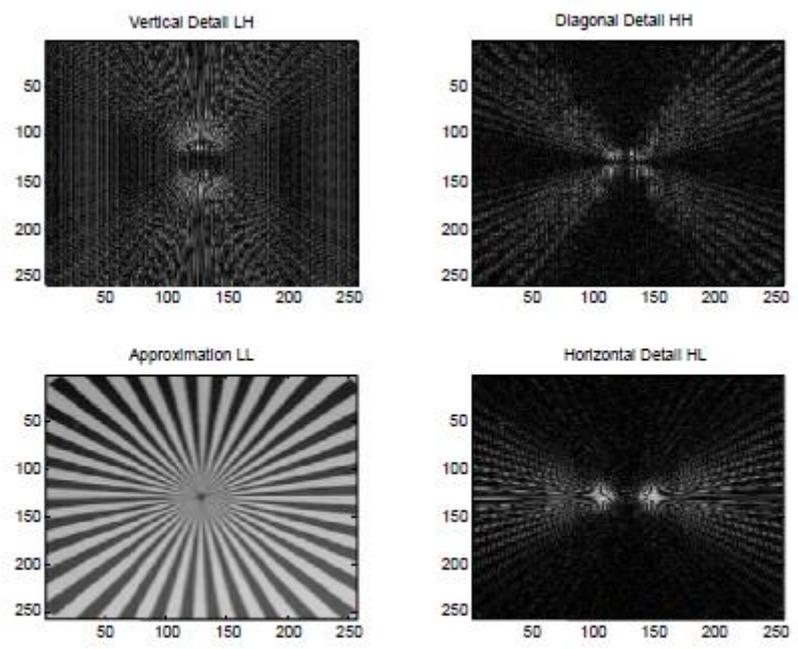


Figure 3. 3: Single level analysis filter bank for DWT.



(a)



(b)

Figure 3. 4: (a) test image, (b) single level DWT decomposition of the test image.

An important property of DWT is perfect reconstruction, or in the other word, the proper retrieval of image. This process is performed by Inverse DWT (IDWT). The analysis filters decomposes the image into different sub-images and the synthesis filters composes the sub-images in order to reconstruct the input image.

Figure 3.5 shows a multilevel decomposition of an image which is achieved by performing the same process on the generated LL image.

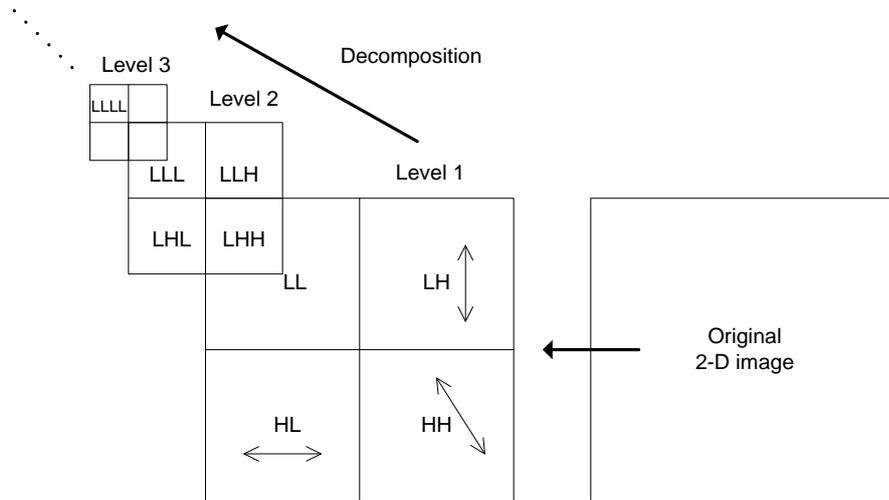


Figure 3. 5: A multilevel decomposition of an image using DWT.

The size of a parent image is four times larger than the size of the child images after decomposition.

An application of DWT is in enhancing the resolution of images which is explained in details in subsection 3.2.3.

3.2.2 Detection of motion and static region

Using pixel subtraction operator two input images produces a third image as an output where, pixel values are simply those of the first image minus the corresponding pixel values from the second image. For instance, considering 2 consecutive images $I_n(x,y)$ and $I_{n-1}(x,y)$, the difference image $I_d(x,y)$ is produced by

$$I_d(x,y) = I_n(x,y) - I_{n-1}(x,y) \quad (17)$$

Moving objects are recognized by an image subtraction algorithm in this method. The major advantage of the algorithm is its simplicity where the implementation is possible in real-time processing of the image processing board, since it simply compares the previous frame with the current one.

If two images have the same pixel value the result of subtraction will be a zero matrix, otherwise, the pixel have a value else than zero, which shows the displacement in the frames. In the video sequences with local motions, subtraction is a simple way to find the motions in sequences.

In this multi frame SR implementation, super resolution is achieved by registering 4 frames. This means all 4 frame differences are required. One way to have these differences is as follows:

- Subtracting reference image (the image that we want to super resolved) from the other 3 input images.
- Applying thresholding for each subtracted image ($I_d(x,y)$) as follows:

For each pixel of $I_d(x,y)$, if the value of the pixel is;

$$I_d(x, y) = \begin{cases} 1 & , \text{if } \mu - 2\sigma \leq I_d(x, y) \leq \mu + 2\sigma \\ 0 & , \text{otherwise} \end{cases} \quad (18)$$

where μ and σ are mean and standard deviation of the pixels in $I_d(x,y)$.

Thresolding is applied to remove the noises in the subtracted images. In this way, more than 90% of the data is thresholded that contains motion regions with less noise.

- OR them up to have all the differences.

Connected component labeling followed by dilation produces the moving region.

3.2.3 Super Resolution Process

Each moving object is taken as an individual image that will be super resolved separately. Using DWT each motion region of input frames is divided into 4 frequency subbands of LL, LH, HL and HH. In this method, Irani et al. super resolution method is applied to each subband motion regions of video sequences separately. This process results in 4 resolution enhanced subbands. Next step is to combine the resolution enhanced subbands using IDWT to produce the super resolved motion region. The resulting super resolved motion region contains sharper edges. This is due to the fact that, the super resolution of isolated high frequency components in HH, HL and LH preserves more high frequency components after the super resolution of the respective subbands separately than super resolving the low resolution image directly. Also the local registration of the motion region results in a more accurate registration process which improves the quality of the super resolved frame in comparison to the classical super resolution with global registration. In parallel the static region is also transformed into wavelet domain for further processing. Bicubic interpolation takes the place of Irani et al. reconstruction and a similar process is applied to the subbands of static region, in order to obtain the resolution enhanced static region. Finally, super resolved motion regions are combined with interpolated static region in order to generate the final super resolved frame.

3.3 Motion-Block-based Localized Super Resolution using Complex Wavelet Transforms (MBSR)

This section gives the details explanation about the super resolution technique discussed in [31]. In this method (MBSR) after detecting the motion region, it is

divided into small blocks in order to increase the possibility of having one direction in each block. DT-CWT is used to decompose the motion and static blocks of the frame into different subbands. The acquired subbands are processed separately and IDT-CWT is used to compose them back and form the super resolved block.

Any classical multi-frame super resolution can be used in this method. The selected registration algorithms used in this work for comparison purposes are:

- Marcel et al.[53],
- Vandewalle et al. [54],
- Keren et al. [3],

The reconstruction methods following the registration process are listed below:

- Interpolation
- Iterated Back Projection (IBP) [4],
- Robust super resolution technique [67],
- Structure Adaptive Normalized Convolution (SANC) [55].

3.3.1 Dual Tree Complex Wavelet Transform (DT-CWT)

The ordinary decimated DWT is shift variant due to the decimation operation exploited in the transform. So that, a small shift in the input signal can results in a very different set of wavelet coefficients. Kingsbury [68] introduced decompositions that not only remove the above problem but also, improve the directional selectivity and perfect reconstruction in comparison to decimated DWT. This wavelet transform is called dual-tree complex wavelet transform (DT-CWT) [69-71] which includes two parallel real filtered wavelet trees with the wavelets forming (approximate) Hilbert pairs. One tree generates the real part of the transform while the other is used in generating complex part [72]. As shown in figure 3.6, $\{h_0(z), h_1(z)\}$ is a Quadrature Mirror Filter

(QMF) pair in the real-coefficient analysis branch. For the complex part, $\{g_0(z), g_1(z)\}$ is another QMF pair in the analysis branch.

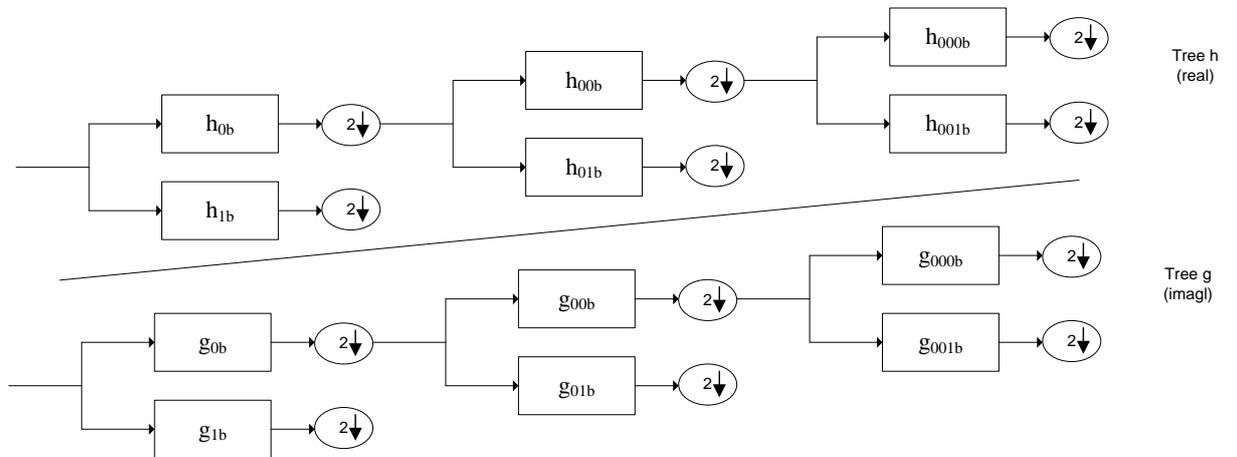
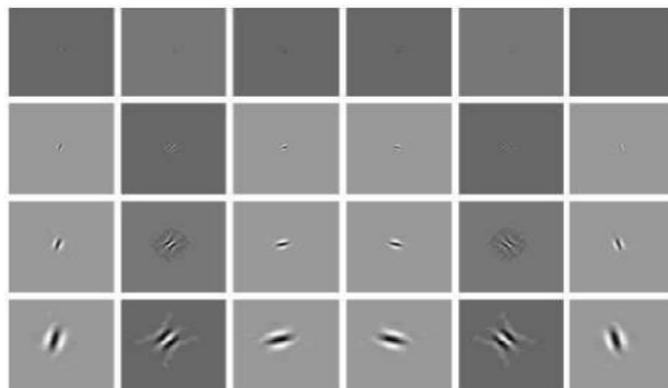


Figure 3. 6: Block diagram for a 3-level DT-CWT [72].

All filter pairs are orthogonal and real-valued. It has been shown [72] that if filters in both trees are made to be offset by half-sample, two wavelets satisfy Hilbert transform pair condition and an approximately analytic wavelet is given by

$$\psi(t) = \psi_h(t) + j\psi_g(t) \quad (19)$$



(a)

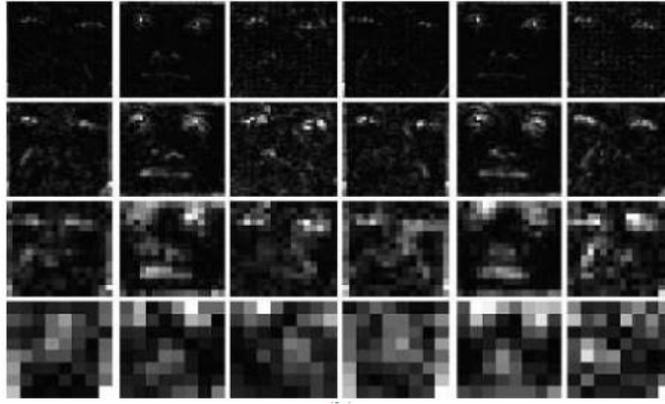


(b)

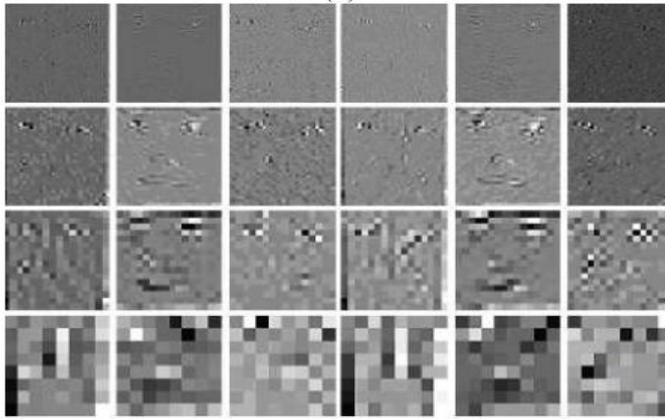
Figure 3. 7: Impulse response of dual-tree complex wavelets at 4 levels and 6 directions. (a) Real part. (b) Magnitude.



(a)



(b)



(c)

Figure 3. 8:(a) Sample image for transformation. (b) The magnitude of the transformation. (c) The real part of the transformation [73].

where $\psi_h(t)$ and $\psi_g(t)$ are two real discrete wavelet transforms employed in parallel to generate the real and imaginary parts of complex wavelet $\psi(t)$.

It has the ability to differentiate positive and negative frequencies and produces six subbands oriented in $\pm 15^\circ, \pm 45^\circ, \pm 75^\circ$.

Figure 3.7 shows the impulse responses of the dual-tree complex wavelets. It is evident that the transform is selective in 6 directions in all of the scales except the first.

Figure 3.8 shows the magnitude and real part of a face image processed using the DT-CWT [73].

In this work DT-CWT is chosen due to its strength in directional selectivity. High frequency details in six different directions are isolated in different subbands and processed separately. This approach helps to minimize the effect of one directional high frequency component over the other directional high frequency component through the super resolution process.

Restoring the high frequencies of a low resolution image is a key to improve its resolution. Various interpolation methods and wavelet transforms are used to solve this problem in [46, 47]. These methods try to recover the original image by processing a low resolution image. The consecutive frames in a video sequence usually contain the same point of view with a small difference. Combining the information of these frames results in a frame with higher resolution.

3.3.2 Motion detection

The principle of motion detection algorithm is to generate a reliable background model and thus significantly improve the detection of moving objects. The three major classes of methods for motion detection are background subtraction, temporal differencing, and optical flow [75]. A recent state-of-the-art motion detection algorithm involving three modules: a background modeling (BM) module, an alarm trigger (AT) module and an object extraction (OE) module is used in this work [32]. The block diagram of the motion detection method is shown in Figure 3.9.

For the BM module, a unique two-phase background matching procedure is performed using rapid matching followed by accurate matching in order to produce optimum background pixels for the background model.

At first, for each pixel (x, y) , the modified moving average (MMA) is used to compute the average of frames 1 through K for the initial current background model, $B_t(x, y)$ generation.

Optimum background modeling is performed by rapid matching to determine the candidates for the next stage which is stable signal trainer. This is accomplished by verifying whether or not the respective pixel values for the incoming video frame $I_t(x, y)$ are equal to the corresponding pixel values of the previous video frame $I_{t-1}(x, y)$.

The candidate pixels then pass through the stable signal trainer as follows:

$$M_t(x, y) = \begin{cases} M_t(x, y) + p, & \text{if } I_t(x, y) > M_{t-1}(x, y) \\ M_t(x, y) - p, & \text{if } I_t(x, y) < M_{t-1}(x, y) \end{cases} \quad (20)$$

The initial background candidate value $M_0(x, y)$ is set at $I_0(x, y)$ where, $M_t(x, y)$ is the corresponding pixel within the most recent set of background candidates, $M_{t-1}(x, y)$ is the corresponding pixel within the previous set of background candidates, and p represents the real value which is experimentally set at 1. Accurate matching procedure obtains the optimum background pixels when the pixels of $M_t(x, y)$ are equal to $I_t(x, y)$. To smooth the background model a simple moving average method updates it. The absolute difference $\Delta_t(x, y)$ is generated by the absolute differential estimation between the updated background model $B_t(x, y)$ and current incoming video frame $I_t(x, y)$ to be used in the next stage of the motion detection method.

The AT module eliminates the unnecessary examination of the entire background region, allowing the subsequent OE module to only process blocks containing moving objects.

In this part, each $w \times w$ block (i, j) within the $\Delta_t(x, y)$ is composed of V discrete gray-levels. The block-based probability density function $P_h^{(i,j)}$ is defined as follows:

$$P_h^{(i,j)} = n_h^{(i,j)} / w^2 \quad (21)$$

where h represents the arbitrary gray-level within each $w \times w$ block (i, j) and $n_h^{(i,j)}$ denotes the number of pixels corresponding to arbitrary gray-level h . h is reset to 0 when it is smaller than τ , where w and τ are experimentally set at 8 and 10 respectively. Detection of each possible motion block candidate is accomplished by the block-based entropy evaluation ($E(i, j)$).

Consequently, after each $w \times w$ entropy block $E(i, j)$ is calculated and the motion block $A(i, j)$ containing pixels of moving objects is labeled with “1,” where the static ones are labeled with “0.”

Elimination of some of the detected background blocks and completed motion blocks is then performed via the block-based morphological erosion and dilation operations shown below:

$$A^* = \delta_{b_{2\lambda}}(\varepsilon_{b_\lambda}(A)) \quad (22)$$

where ε is the morphological erosion, δ is the morphological dilation, and b_λ is a ball of radius λ which acts as the elemental structure and experimentally is set at 1.

The OE module forms the binary object detection mask in order to achieve highly complete detection of moving objects.

A threshold selection algorithm for use with the OE module in order to produce the binary motion detection mask obtains a suitable threshold for binarization.

A variance value $V_t(x, y)$ is calculated at each frame and finally, the binary motion detection mask $D(x, y)$ is formed by detecting the pixels of moving objects within each motion block as follows:

$$D(x, y) = \begin{cases} 1, & \text{if } \Delta_t(x, y) > V_t(x, y) \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

The input to the motion detection method is a video. A multi-frame super resolution method is performed in this method which uses four consecutive frames. Therefore, each group of four consecutive frames is utilized as a sequence to be used as the input video to the motion detection method. The current frame of each of these video sequences is the frame to be super resolved (reference frame). As a result the total motion regions of the frames with respect to the reference frame can be extracted.

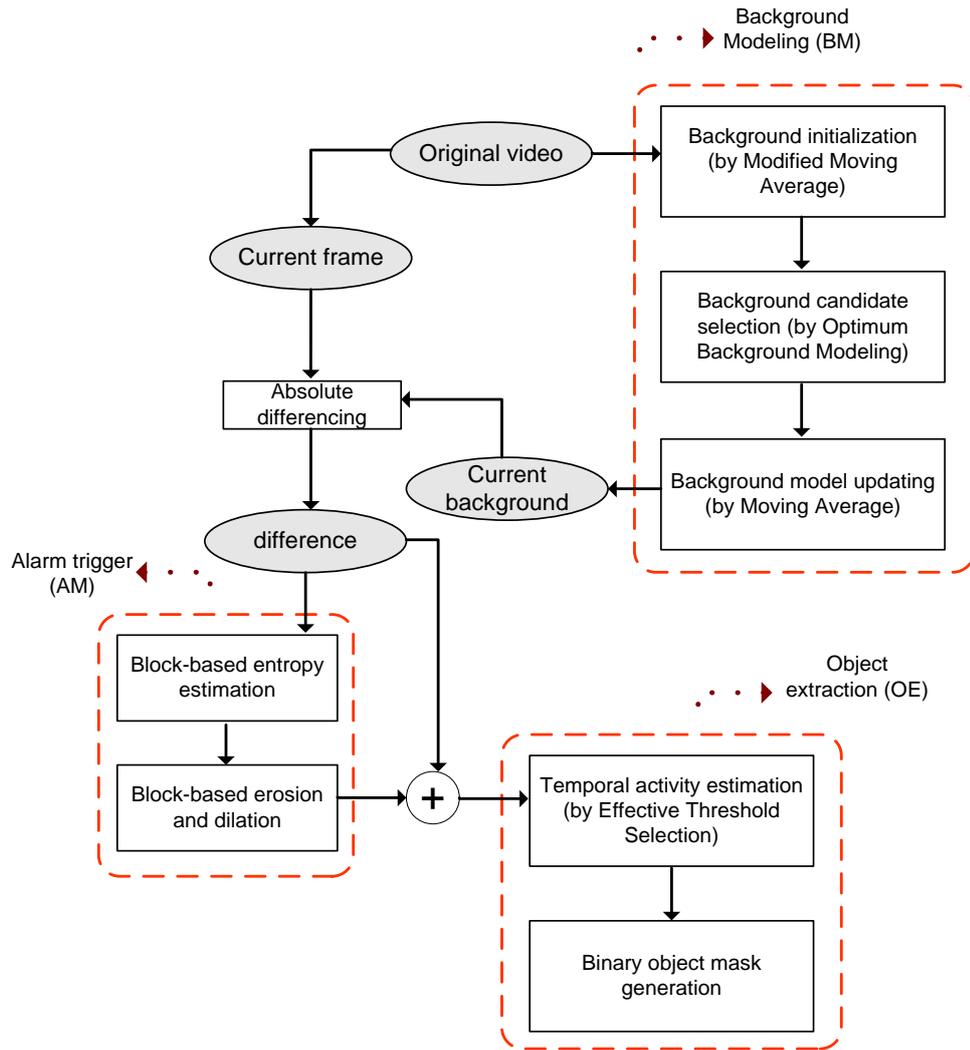


Figure 3. 9: Block diagram of the motion detection method.

3.3.3 Extraction and insertion of the motion blocks

Once the motion regions are detected, they should be extracted in blocks for further processing. MBSR method separates the motion regions from the background and extracts the motion regions into multiple 32x32 motion blocks. These blocks will be treated as a low resolution image to be super resolved via a multi-frame super resolution algorithm and then to be inserted back to the background. However, the main drawback of SR algorithms is the generated artifacts around the boundary of the super

resolved image. These artifacts cause degradation in the quality of the final result. This degradation is within a 4-row pixels distance of the boundary. In order to overcome this problem, the central 28x28 pixels region (see figure 3.10) of the extracted 32x32 pixels is inserted into the final image. Figure 3.11 shows the block diagram of the motion block extraction algorithm.

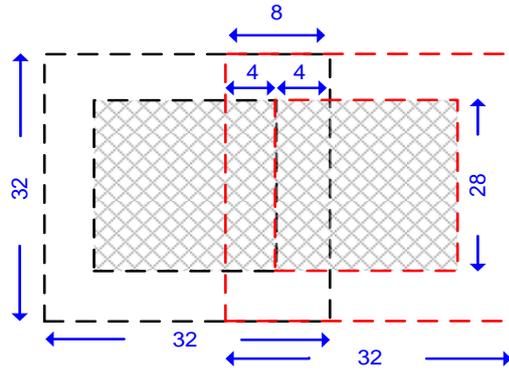


Figure 3. 10: Two neighboring motion blocks. Measures are in pixels.

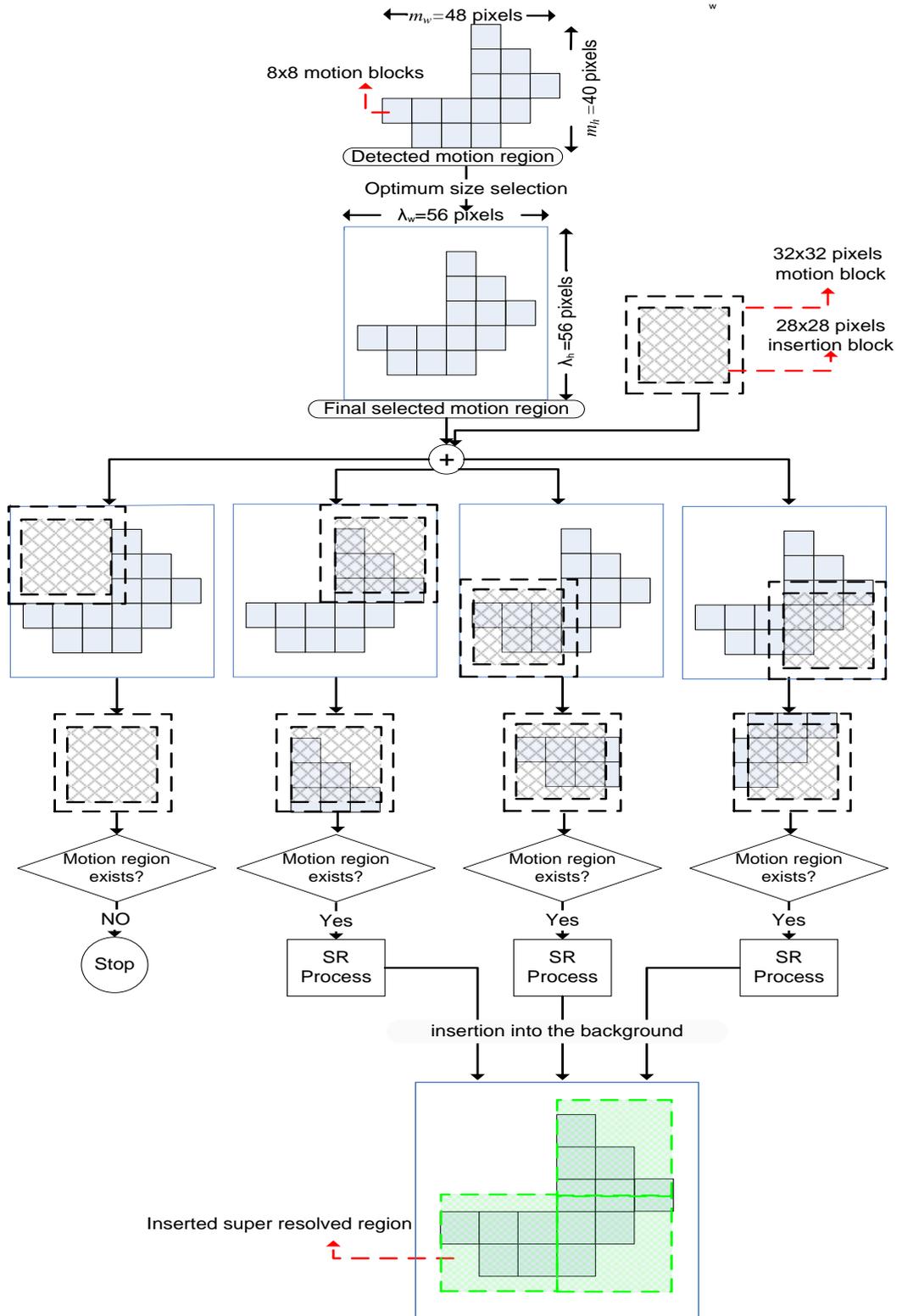


Figure 3. 11: Block diagram of the motion block extraction.

Consider the “detected motion region” shown in figure 3.11. As the motion detection algorithm dictates, this region consists of 8x8 motion blocks. In this example, the maximum width (m_w) of the region is 48 pixels, where the maximum height (m_h) of it is 40 pixels. The next stage is to find a region in which we can position an integer number of 32x32 pixel blocks. For this purpose a formula is defined as follows:

$$\lambda_w = 24\beta_w + 8 \quad (24)$$

$$\lambda_h = 24\beta_h + 8 \quad (25)$$

where β_w and β_h are two positive integer numbers and λ_w and λ_h are the width and the height of the “final selected region” respectively ($m_w \leq \lambda_w$ and $m_h \leq \lambda_h$). In figure 3.11 the size of “detected motion region” is 40x48 pixels which are extended to 56x56 pixels for the “final selected motion region”. β_w and β_h are both calculated as 2, which means we can place two rows and two columns of 32x32 pixel blocks over the “final selected motion region”. At this point, the algorithm checks the existence of the motion region in the central 28x28 pixels block of each placed 32x32 pixel blocks. If motion region exists the 32x32 pixels block will be sent to the SR process, otherwise it is ignored. After the SR process the size of each 32x32 motion blocks will be $\alpha 32 \times \alpha 32$. The central $\alpha 28 \times \alpha 28$ part of the super resolved motion block is inserted into the background, which is free of boundary artifacts.

3.3.4 Super Resolution Process

Figure 3.12 illustrates the complete procedure of the resolution enhancement technique. A one-level DT-CWT decomposes the motion blocks and remaining static background to complex-valued low and high frequency subbands with more directivity

in comparison to DWT [64, 74]. Then six complex-value high frequency subbands with $+75^\circ$, $+45^\circ$, $+15^\circ$, -15° , -45° , and -75° orientations of the respective stationary part have been enlarged by using bicubic interpolation. In parallel a super resolution method is applied to the same subbands related to small region. Having found the enlarged subbands of static and motion regions, we can compose them back by exerting Inverse DT-CWT. At this point, super resolved motion and stationary regions are ready to be combined and form the result of this resolution enhancement technique.

Step by step description of MBSR method is given below:

Step 1: Acquire frames from video and detect motion region(s) by motion detection algorithm [32].

Step 2: Extract the 32x32 motion blocks using the developed motion block extraction algorithm.

Step 3: Apply CT-DWT to decompose the motion blocks into different subbands

Step 4: Super resolve extracted motion blocks obtained in step 2.

Step 5: Apply bicubic interpolation to the high frequency subbands of each motion blocks separately.

Step 6: Generate super resolved motion blocks by using ICT-DWT for low frequency subband generated in step 4 and high frequency subbands obtained from step 5.

Step 7: Apply CT-DWT to decompose the background region into different subbands.

Step 8: Apply bicubic interpolation to the entire static region and to the high frequency components obtained in step 7.

Step 9: Generate interpolated background by using ICT-DWT for the subbands generated in step 8.

Step 10: Inserting the super resolved motion blocks obtained from step 6 into the interpolated background attained in step 9.

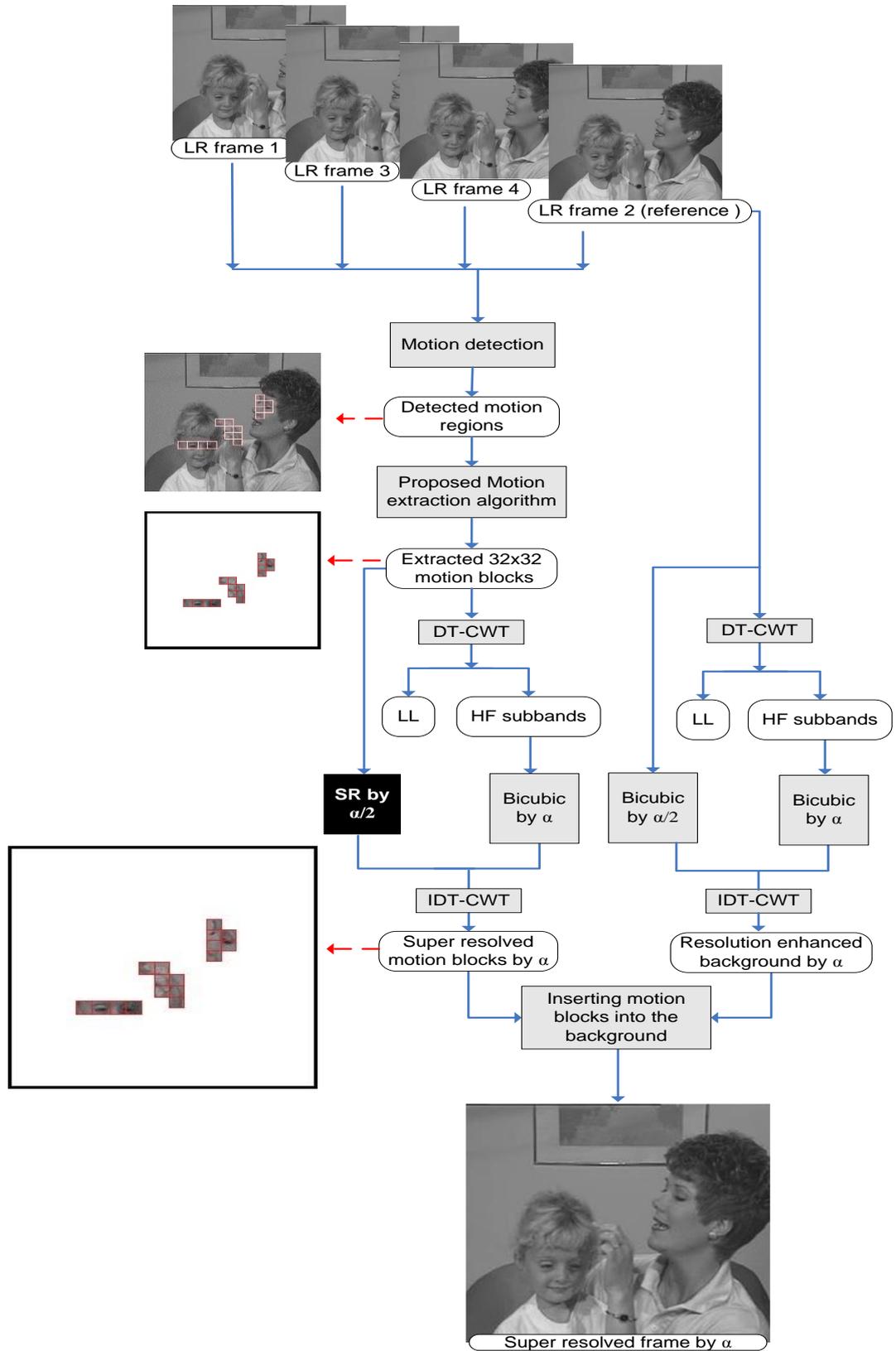


Figure 3. 12: The block diagram of the MBSR method presented in [31].

3.4 Motion-block-based Localized Super Resolution using New Edge Directed Interpolation and Complex Wavelet Transforms (MBSR using NEDI)

The idea of super resolution is based on the theory that there exists a mathematical model which could have generated the LR observation from the unknown high-resolution (HR) image. The goal of every SR method is to resolve this mathematical model; nevertheless, because of the inaccuracies in estimation of its parameters an erroneous model may be obtained. These errors, even small in number, can result in an undesirable image reconstruction. In such a case interpolating the reference frame, results in achieving a higher resolution image in comparison to an image created using incorrect motion parameters [54]. Robust procedures are necessary to provide reliable estimates. Cheng et al. [37] presented an interesting SR approach for video sequences using modified non-local means. They defined the so called complex areas in the LR reference that included non-translation motion, occluded, and new-appearing object areas. They named the remaining part of the frame as simple area which contained the static and near translation motion areas. In their approach each reference frame was divided into multiple patches of the same size. If the patch was located in a simple area, it could be easily super resolved using various similar patches. The problem occurred when there were few sub-similar patches related to the patches in the complex area. This caused degradation in the visual quality of the final video SR. To solve this issue, they reduced the size of the patch so that the probability of finding similar patches would increase. Due to their priorities to reduce the computational complexity of this method, the visual quality of the reported results has limited performance increase.

In this technique dividing every frame into stationary and motion parts helps in determining proper registration. However, in the motion part of an image there may be different activities in different directions which cause incorrect motion estimation. Similar to [37], dividing the motion region of the image into small blocks is the solution of this method to this problem. Increasing the probability of having dedicated direction of shift and rotation in each block, results in obtaining less error in the registration of images. These blocks can then be treated as if taken from different low resolution images for registration purposes.

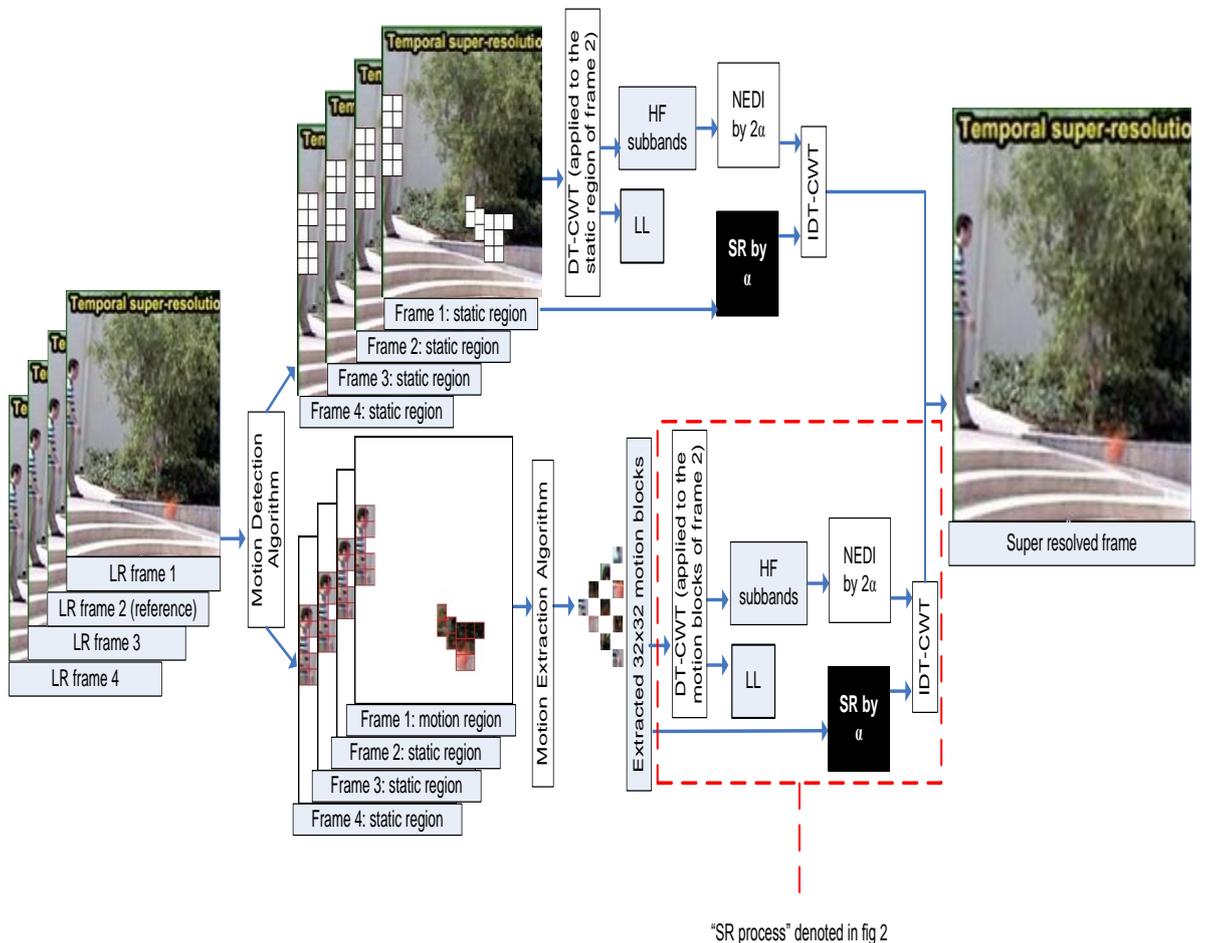


Figure 3. 13: Block diagram of the MBSR using NEDI technique [31].

Bicubic interpolation is one of the well-known interpolation methods, which has been used in different wavelet based image resolution enhancement methods [76]. However, blurred edges or artifacts around edge area are the results of smoothing property of the bicubic interpolation. As many SR algorithms [36, 40, 48] tried to work out this issue, edge-directed interpolation (EDI) [48] is used as an alternative interpolation method in wavelet based image resolution enhancement method [77]. In this work we applied new edge-directed interpolation (NEDI) [40] which improved the performance of EDI [78] by obtaining a resolution enhanced image which is not smooth perpendicular to edges and is smooth parallel to edges.

Figure 3.13 illustrates the complete procedure of this resolution enhancement technique.

The stages of the technique demonstrated in [31] are outlined below:

Step 1: Acquire frames from the low resolution video and detect motion region(s) by motion detection algorithm [32].

Step 2: Extract the 32x32 motion blocks using motion block extraction algorithm. Employ DT-CWT to decompose the low resolution motion blocks into Low and high frequency subbands.

Step 3: Super resolve extracted motion blocks of step 2 using classical SR methods.

Step 4: Apply NEDI to the high frequency subbands (with $+75^\circ$, $+45^\circ$, $+15^\circ$, -15° , -45° , and -75° orientations) of each motion blocks separately.

Step 5: Generate super resolved motion blocks by using IDT-CWT for low frequency subband generated in step 3 and high frequency subbands obtained from step 4.

Step 6: Super resolve the static region using classical SR methods.

Step 7: Apply DT-CWT to decompose the background region into different subbands.

Step 8: Use NEDI to the high frequency components obtained in step 7.

Step 9: Generate the super resolved background by using IDT-CWT for the subbands generated in step 8 and 6.

Step 10: Insert the super resolved motion blocks obtained from step 5 into the interpolated background attained in step 9.

Occlusion refers to covering or uncovering a part of an object or scene due to object or camera motion which causes problem in determining the HR image. To deal with this problem, Alvarez et al. [79] labeled each pixel of the LR image as “Observable” and “unobservable”. In this method the unobservable pixels were assumed to be constants and the observable pixels were used in the SR reconstruction. Shen et al.[80] applied the same terms to denote the type of pixels. They defined a threshold for a more appropriate approximation of the unobservable pixels. Consequently, the artifacts were effectively suppressed around motion boundaries and occlusion regions when compared with [79].

Occlusion is one of the problems, which is not directly addressed in this work. However, the idea of using multiple consecutive low resolution frames to generate the

high resolution super resolved frame, indirectly incorporates the occluded pixels in some of the LR frames into the output HR frame.

3.5 Simulation Results and Discussions

The MBSR using NEDI technique is tested on three well known benchmark video sequences such as “Akiyo”, “Mother & daughter” and “Container”. The original high resolution benchmark video sequences have the size of 512x512 pixels per frame which are reserved as the ground truth for evaluating the performance of the enhancement quality in terms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). The input low resolution frames with 128x128 pixels are generated such that, each ground truth frame is blurred by a Gaussian low-pass filter with the mask size of 4x4 and standard deviation of 1. Then, the blurred frames are down-sampled vertically and horizontally with a factor of 4 and Gaussian noise was added to down-sampled frames.

One of the most important parts of the MBSR using NEDI technique is the super resolution process (denoted by black box in figure 3.13) applied for the static region as well as the blocks extracted from the motion regions. In this work, a state of art registration (Vandewalle et al. [54]) and reconstruction ((SANC) [55]) methods are selected to be used in this technique.

In the first experiment, as a preliminary step towards creating an algorithm which can deal with each motion direction individually, we perform a block size searching. The resulting average PSNR and SSIM performances of the MBSR using NEDI technique using different block sizes have been shown in Table 3.1. The results indicate that, in all three video sequences the highest performance is obtained for the 32x32 pixels block size. Large block sizes such as 64x64 decreases the local registration accuracy. On the other hand the small block size such as 16x16 lacks the required information content and hence generates relatively lower PSNR and SSIM

performance. Based on the results, 32x32 block size has been utilized in the MBSR using NEDI technique.

Table 3. 1: Average PSNR (dB) and SSIM values obtained after applying different motion block sizes in the MBSR using NEDI technique.

Block Size	Mother & daughter		Akiyo		Container	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
128X128	31.73	0.9223	30.01	0.9380	27.83	0.8401
64X64	33.84	0.9462	32.98	0.9498	29.32	0.8763
32X32	34.02	0.9458	33.76	0.9531	29.54	0.8775
16X16	33.60	0.9461	33.01	0.9505	29.30	0.8770

The focus of this work is on two issues, obtaining the best size of the motion block to be used in SR so that registration parameters can be determined properly and employing an interpolation method which can possibly preserve the edges. Next experiment considers the optimal motion block size and different interpolation methods to be used in this technique. The experimental results are given in Table 3.2. This table is organized in a way that it specifies the improvements introduced by each part of the proposed algorithms. In the Classical SR each full frame is processed through Vandewalle registration followed by SANC reconstruction and the resulting average PSNR and SSIM is reported for each video. As mentioned before MSR method detects the motion region and uses DWT to improve the resolution of each frame. To find out the improvements achieved by each part of this method, in the next experiment motion region is detected and each rectangular motion region is super resolved separately. The

static background is resolution enhanced using bicubic interpolation. In fact, DWT is removed from MSR to show the improvement achieved by just separating the rectangular motion regions from the static background. Adding DWT to the experiment shows the effect of DWT on the MSR method. To show the disadvantage of not dividing the motion region into 32x32 blocks and leaving the motion part as a single rectangular region, motion based methods are utilized. Bicubic interpolation vs NEDI are used to show the effectiveness of NEDI on preserving the edge quality. Two different motion region detections are introduced in sections . Both methods are applied on MBSR and the results show that method comes up with a slightly better SSIM and PSNR. DWT is applied on MBSR in another experiment to specify the improvement achieved by DT-CWT in comparison with DWT. In all cases the MBSR using NEDI technique achieves better results in comparison to the classical SR methods.

The performance gain achieved by the MBSR using NEDI technique can be attributed primarily to the localization of motion blocks going through a dedicated SR process. On the other hand, DT-CWT helps to decompose the given motion and static region images into different high frequency directional subbands which isolates the edge details in different directions and hence reduces the undesirable inter-directional interference in the SR process.

In order to show the superiority of the MBSR using NEDI technique for each frame we have devised a separate experiment. The PSNR of each resolution enhanced frame in “Mother & daughter” video sequence is plotted in figure 3.14 for classical SR method versus MBSR using NEDI technique. Note that the MBSR using NEDI technique outperforms the classical SR for every frame in the video sequence.

Table 3. 2: The average PSNR (dB) and SSIM values of different resolution enhancement methods on the test video sequences.

RESOLUTION ENHANCEMENT METHODS	Mother & daughter		Akiyo		Container	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic Interpolation	27.44	0.8884	27.20	0.9152	23.12	0.8111
NEDI [48]	28.78	0.8959	28.01	0.9263	24.71	0.8233
Classical SR [54,55]	30.25	0.9148	29.30	0.9312	26.82	0.8226
MSR without DWT	31.45	0.9221	30.56	0.9367	27.82	0.83.44
MSR [30]	32.12	0.9324	31.82	0.9404	28.19	0.8422
MSR using NEDI	32.65	0.9348	32.35	0.9489	28.71	0.8427
MBSR using Frame differences	33.12	0.9439	32.88	0.9504	28.93	0.8710
MBSR using DWT	32.83	0.9412	32.65	0.9497	28.78	0.8672
MBSR	33.22	0.9441	32.92	0.9505	28.98	0.8713
MBSR using NEDI	34.02	0.9458	33.76	0.9531	29.54	0.8775

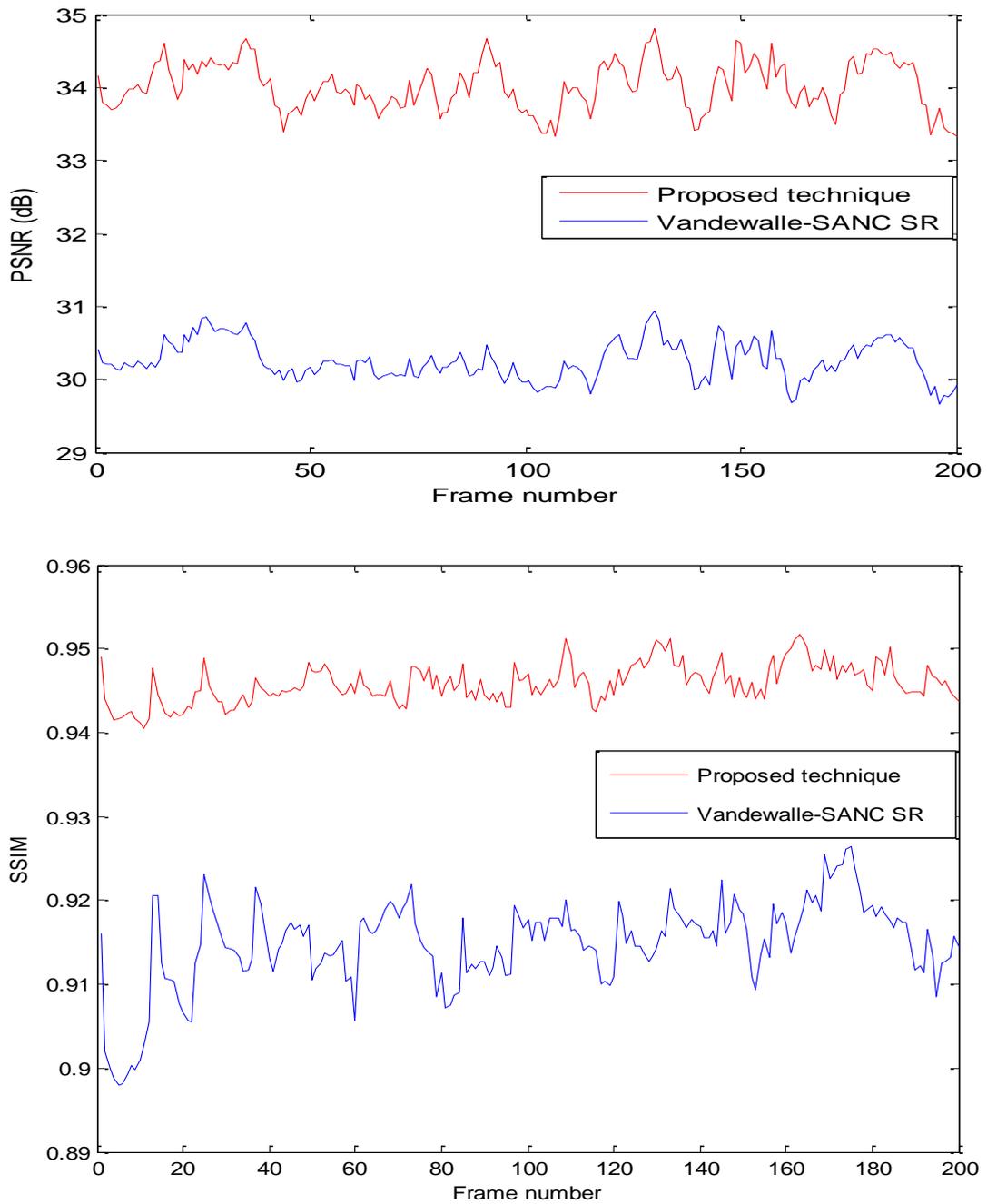


Figure 3. 14: PSNR and SSIM result of resolution enhancement of “Mother & daughter” video sequence obtained from Vandewalle and SANC SR versus MBSR using NEDI technique for 200 consecutive frames.



Frame 59: a

b

c (30.12, 0.9108)

d (33.89, 0.9458)



Frame 80: a

b

c (30.14, 0.9114)

d (33.75, 0.9444)



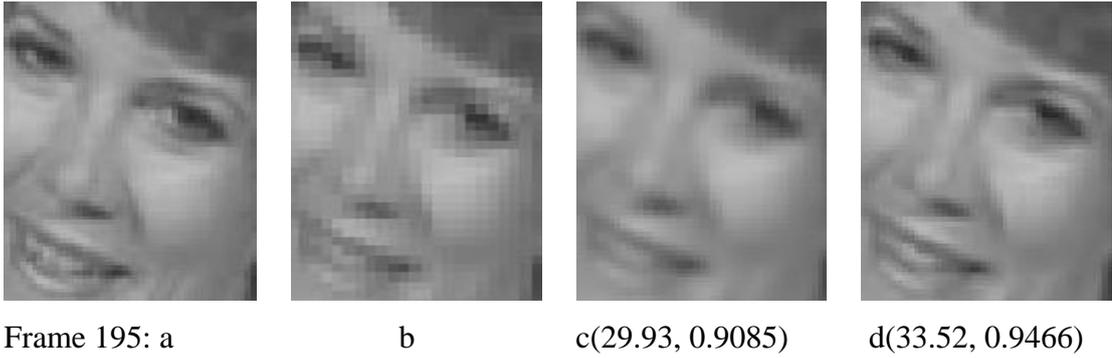
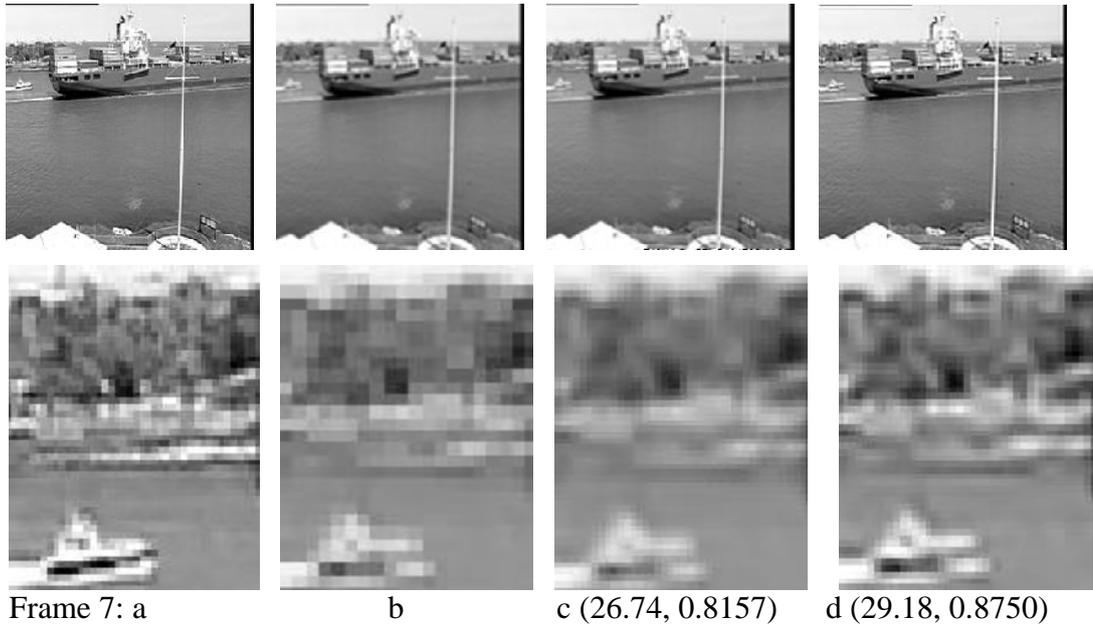


Figure 3. 15: Result of different SR methods on "Mother & daughter" video frames (PSNR, SSIM in parenthesis),(a) Reference HR frame. (b) Input LR frame. (c) SR using [54] [55] (d) MBSR using NEDI technique.



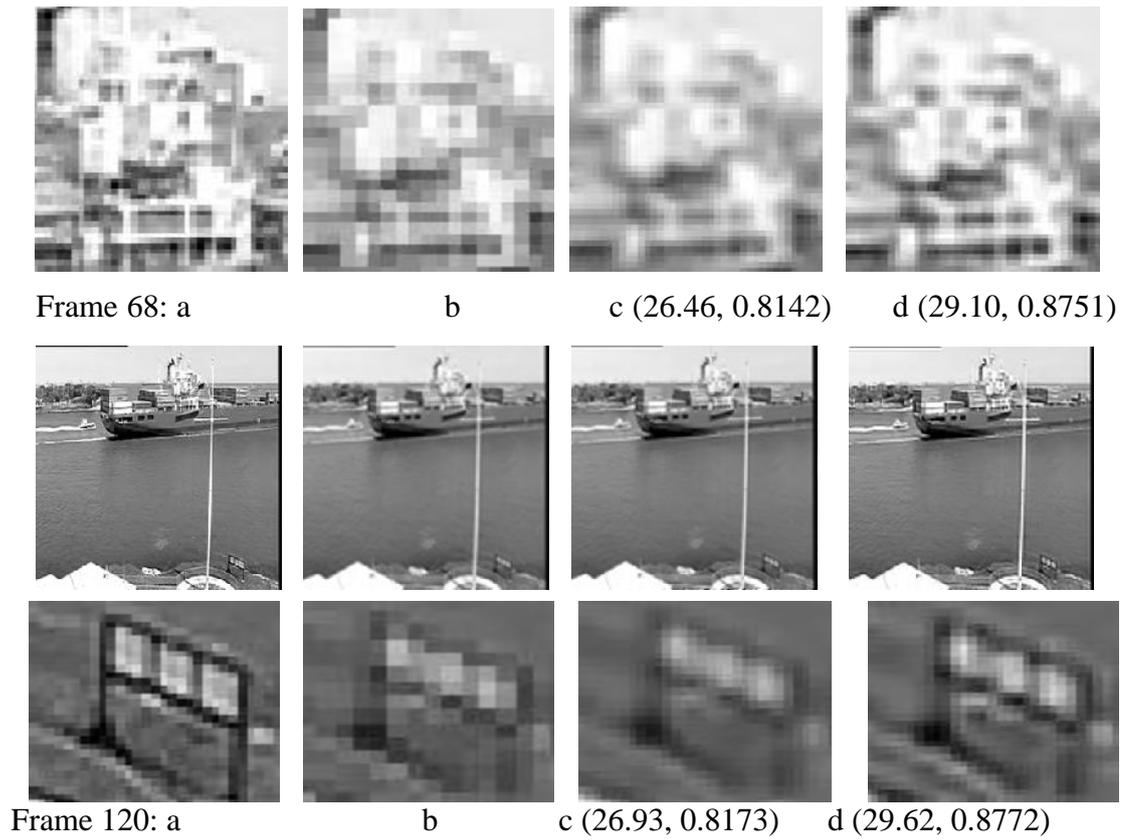


Figure 3. 16: Result of different SR methods on “Container” video frames (PSNR, SSIM in parenthesis), (a) Reference HR frame. (b) Input LR frame. (c) SR using [54] [55] (d) MBSR using NEDI technique.



comparison to the classical SR is based on the fact that we are separating the motion and the static regions to be treated differently. The motion parts benefit from the calculated motion parameters in the registration process. The block based approach also provides more accurate registration of the consecutive frames. This is simply because localized registration is more efficient than global registration of the entire frame. Furthermore, using DT-CWT based decomposition with 6 directions preserves the integrity of the high frequency components such as edges throughout the this SR process.

The performance of the MBSR using NEDI technique is compared with the state of the art Generalized NLM-SR presented by Protter et.al[81]. Three real video sequences, “Foreman”, “Miss America” and “Suzie” are used in this experiment. The average PSNR and SSIM results on these test sequences are shown in Table 3.3. As you noticed, MBSR using NEDI technique generates better results in comparison to NL-Means method in most of the cases. Precise motion estimation and using the information of several LR images are the advantage of the MBSR using NEDI technique over the challenging NLM-SR. The PSNR and SSIM of each super resolved frame in “Foreman” sequence is plotted in figure 3.18 for Protter et al. method versus MBSR using NEDI technique. The PSNR and SSIM values are higher for the MBSR using NEDI technique when compared with NLM-SR [81]. Figure 3.19 shows visual results of this comparison. The results show that NLM-SR method produces excessively sharp edges in the form of block effects while the MBSR using NEDI technique does not suffer from this problem.

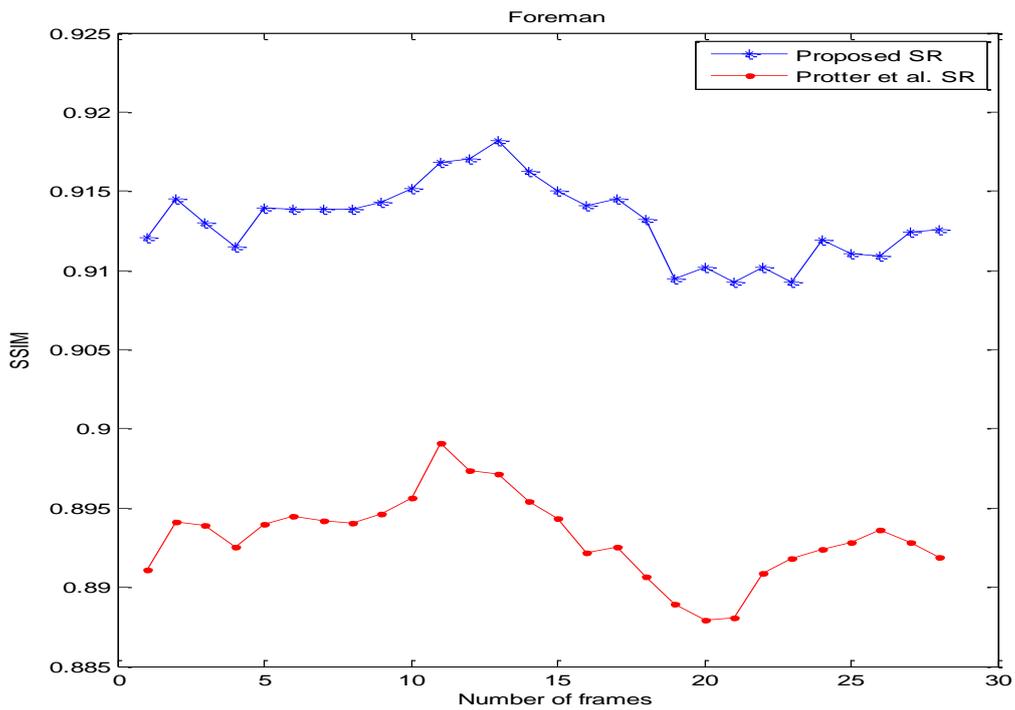
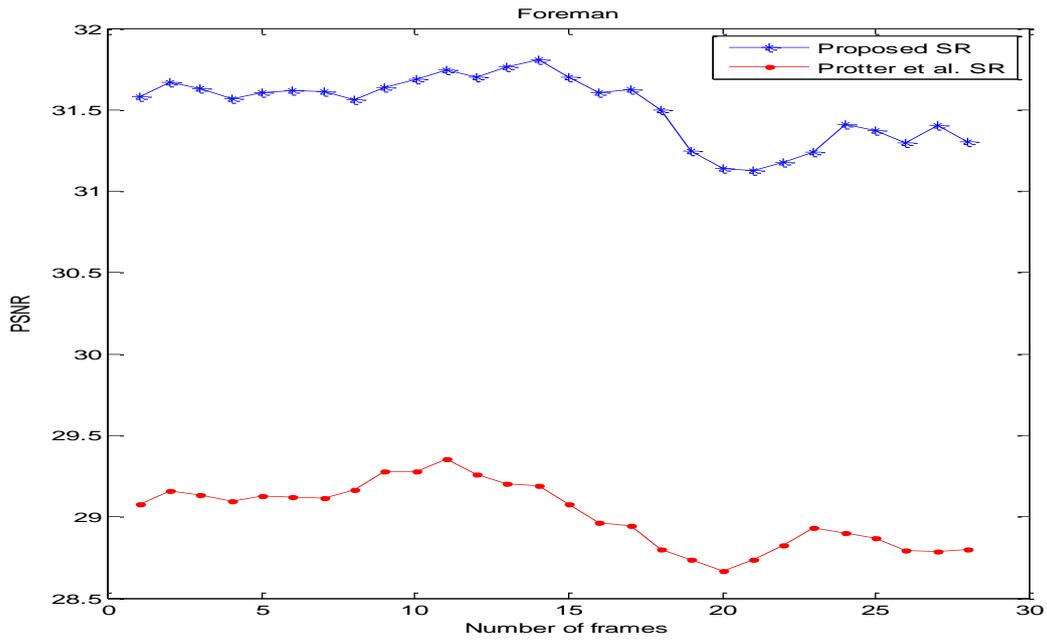


Figure 3. 18: PSNR result of resolution enhancement of “Foreman” video sequence obtained from Protter et al. SR versus MBSR using NEDI technique for various frames.



(a)



(b) (PSNR=29.35 SSIM=0.8991)



(c) (PSNR=31.74 SSIM=0.9168)

Figure 3. 19: (a) Reference HR image (12th frame) of Foreman video sequence. (b) Protter et al. super resolution (c) MBSR using NEDI technique.

Table 3. 3: The average PSNR (dB) and SSIM values of different resolution enhancement methods on the test video sequences.

Video Sequence	Protter et.al [81]		MBSR using NEDI technique	
	SSIM	PSNR	SSIM	PSNR
Foreman	0.8932	29.01	0.9129	31.51
Miss America	0.9165	35.47	0.9261	35.33
Suzie	0.8770	32.60	0.8765	33.01

Another simulation investigates the effect of motion region detection on the registration part of super resolution algorithms. In this simulation, each frame of the synthetic video sequence (with 100 frames) is formed by rotating and translating an image in the middle of a constant background image. The size of frames is 256x256 and the motion region has the size of 64x64. The rotation and translation parameters are known and varying in each frame in comparison to the neighbor frame.



Figure 3. 20: Frame (b) is the translated and rotated version of frame (a).

As expected and shown in Table 3.4, the rotation and translation parameters obtained by different registration methods have less error (in terms of degrees for rotation and pixels for translation) in the case of block based registration. Figure 3.20 shows two

frame of this video sequence. The frame shown in Figure 3.20 (b) is the translated and rotated version of the frame shown in Figure 3.20 (a).

Table 3. 4: Absolute rotational and translational errors for the synthetic test video sequence.

Method	Average Rotation error (10 degrees)	Average horizontal shift error (10 pixels)	Average vertical shift error (10 pixels)
Keren	10.56	3.96	4.17
Block based keren	1.18	1.04	0.83
Marcel	13.25	4.70	6.10
Block based Marcel	2.34	1.40	1.30
Vandewalle	11.24	5.13	5.40
Block based Vandewalle	1.01	1.18	0.87

CHAPTER 4

THE PROPOSED MOTION BLOCK BASED VIDEO SUPER-RESOLUTION

4.1 Introduction

When a number of low resolution images hold slightly different perspectives of the same frame, by employing an accurate global registration method, the obtained total information about the frame surpasses information from any single frame. When camera moves and the scene are stationary, the relative displacements are global since the whole frame is displaced. In this case, the above mentioned registration methods are able to produce precise results. However, if there are individual objects moving within a frame, it is beneficial to identify the local motion which is determining the motion of each object individually. If an object is static in all frames, no extra information can be gained. Note that, wrong registration can result in an image with a lower quality than the input LR image.

To overcome this problem a block based motion estimation technique using optical flow estimation is proposed. Optical flow estimation is a method to approximate the displacement field between two images. Lucas et al.[83] and Horn et al.[84] were the first researchers to introduce algorithms in this research area. Later on, many new concepts were presented for dealing with shortcomings of previous models [85, 86]. Brox et al. [34] developed an algorithm for computing optical flow by combining a

brightness constancy assumption, a gradient constancy assumption, and a discontinuity-preserving spatio-temporal smoothness constraint. In this work Brox et al. optical flow method has been adapted for estimating the motion field between consecutive frames.

This thesis presents a new video super resolution technique which consists of three modules: pre-processing module, motion block processing module and post processing module. At the pre-processing module, using an optical flow estimation method [34] the input LR frame is divided into motion and static blocks. A classical SR is applied on the static blocks which generates the high resolution static region. The obtained motion blocks from the pre-processing module go through a reconstruction process in the initial motion block reconstruction phase of the motion block processing module in order to produce the high resolution motion blocks. Then, the occluded blocks, which deteriorate the accuracy of the registration process, are identified by calculating an adaptive threshold for each block and are replaced with an interpolated motion block, which more accurately characterizes the occluded region. The last module is the post processing module. At this module, since the blocking artifacts often occur at boundaries between the interpolated occluded blocks and the super resolved blocks, a simple de-blocking filter is applied to the block boundaries. Finally, a de-blurring process is applied to generate the output super resolved frame. The de-blurring is achieved by adding the high frequency subbands of the DWT based interpolated LR input frame to the super resolved frame.

In order to evaluate the performance of the proposed technique, four benchmark CIF video sequences, ice, foreman, news and ice are used. The performance of the

proposed technique in terms of PSNR values and visual quality shows the superiority of the technique presented over the alternative methods in the literature.

The rest of this chapter is organized as follows: section 2 presents the proposed SR technique. The experimental results and discussion is provided in section 3. Finally we concluded in section 4.

4.2 Proposed resolution enhancement method

Multiframe SR improves the spatial resolution of the output frames by combining information of a sequence of subpixel displaced low-resolution frames. The first step in a Multiframe SR is motion estimation. The objective of motion estimation in super resolution is to find a prediction of the pixel displacements in an image with respect to a reference image. Using the computed motion estimation the images are being registered for further processing. Accurate image registration is a fundamental stage in all Super-Resolution processes. Various image registration methods [3, 53, 54] produce good results under global translational motion. However, in practical applications, the global motion pattern is rarely found in the real LR inputs. Instead, typically a combination of isolated local displacements characterizes the motion. Therefore, global registration algorithms are not necessarily appropriate for video super resolution tasks. Interpolating the LR image results in a higher resolution image in comparison to a super resolved image with inappropriate image registration.

In this work a new super resolution technique based on the localization of the motion and static regions in consecutive frames is proposed. Block diagram of the proposed super resolution technique is shown in figure 4.1. The method consists of three modules: pre-processing, motion block processing and post processing. The proceeding subsections explain these three modules in detail.

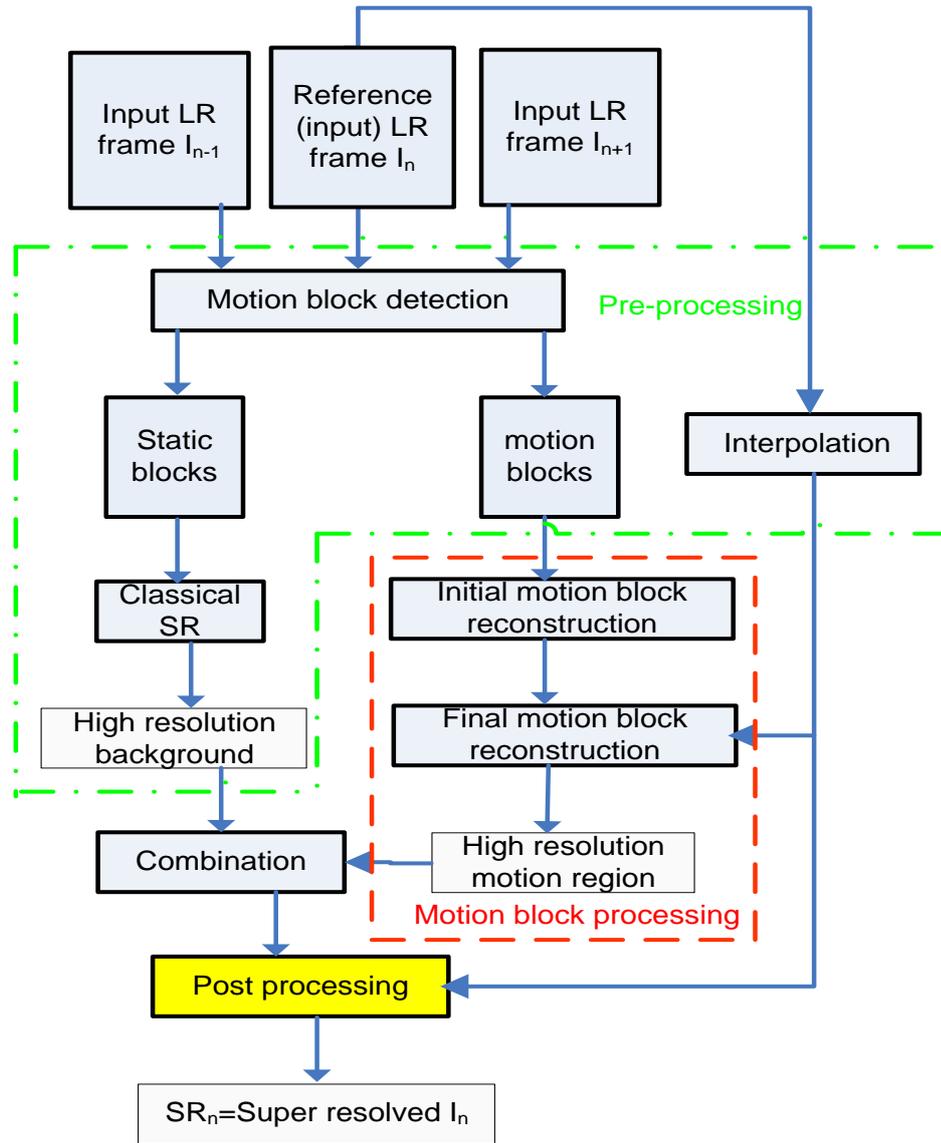


Figure 4. 1: Block diagram of the proposed super resolution technique.

4.2.1 Pre- processing module

The inputs to the proposed technique are three consecutive frames, in which, the middle frame is the reference frame to be super resolved. Let I_{n-1} , I_n and I_{n+1} denote

the three input frames which are a frame before the reference frame, the reference frame and the frame after the reference frame respectively.

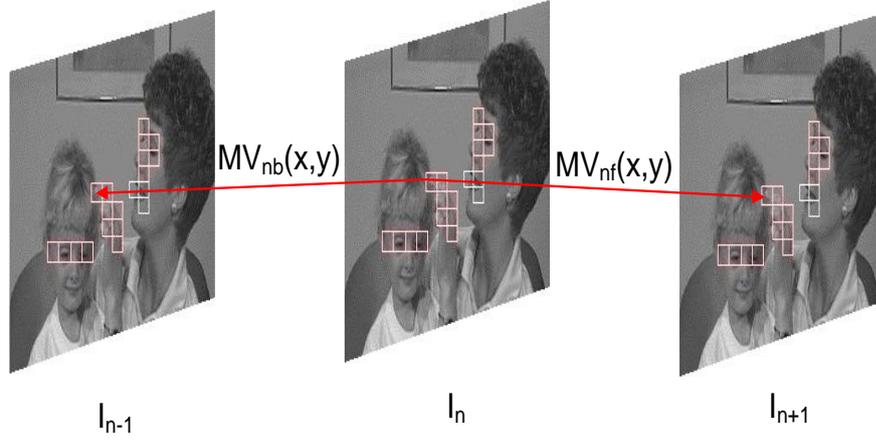


Figure 4. 2: Two motion vectors between three low resolution input frames, corresponding to a motion block at the reference frame.

A more accurate estimation of the local displacements in the consecutive frames is utilized by dividing each frame into blocks of size 8x8 pixels. Matrices MV_{nb} and MV_{nf} contain the estimated motion vectors which are obtained by applying motion estimation [34] on I_n and I_{n-1} , and, I_n and I_{n+1} , respectively. Figure 4.2 illustrates the motion blocks on the three consecutive frames and the motion vectors MV_{nb} and MV_{nf} in one of the 8x8 motion block. Based on the magnitude of the motion vectors assigned to each block, the block is labeled as static or motion block as follows:

Given the three consecutive frames,

$$I_{n-1}(x, y), I_n(x, y), I_{n+1}(x, y)$$

then

$$\begin{aligned}
& \text{if } |MV_{nf}(x, y)| \neq 0 \oplus |MV_{nb}(x, y)| \neq 0 \quad \Rightarrow \text{motionblock}(x, y) \\
& \text{if } |MV_{nf}(x, y)| = 0 \& |MV_{nb}(x, y)| = 0 \quad \Rightarrow \text{staticblock}(x, y)
\end{aligned} \tag{26}$$

where x and y are the row and the column location of the block at the frame respectively. Operators \oplus and $\&$ correspond to logical OR and AND operations respectively. At this point motion blocks are separated from the static blocks. Static blocks are sent to be super resolved and generate the high resolution background, H_{ns} , using one of the state of the art [55] reconstruction methods.

Figure 4.3 shows the detailed block diagram of the pre-processing module. As shown in figure 4.3 an interpolation method is applied to the input reference frame I_n . The resulting image is called H_{nd} which will be used in the other two modules. The reason is clarified in section 5.2.2.

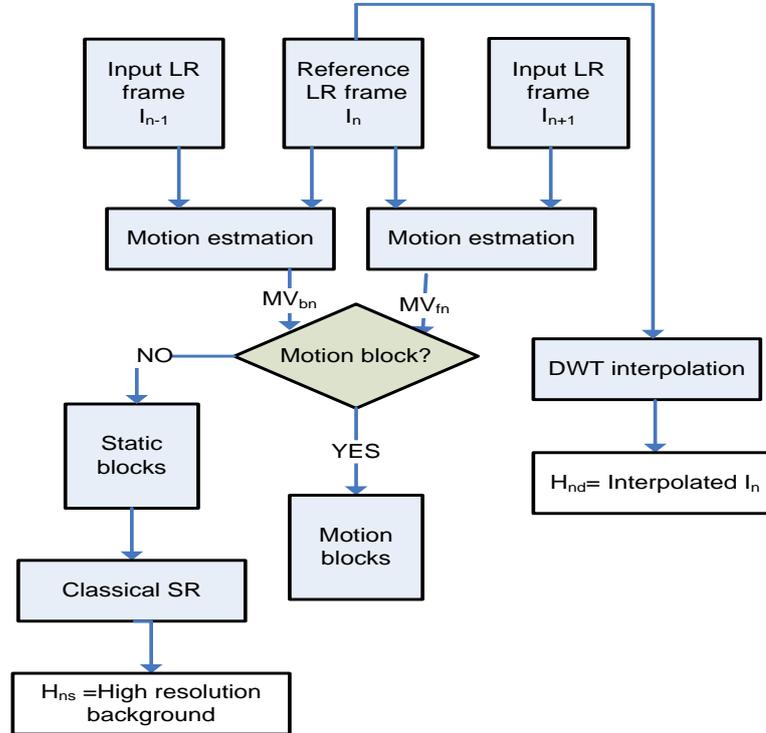


Figure 4. 3: Detailed block diagram of the pre-processing module.

4.2.2 Motion block processing module

4.2.2.1 Initial motion block reconstruction

Here, the focus is to register the motion blocks in an accurate manner. According to MV_{nb} and MV_{nf} each block in the reference frame $I_n(x,y)$ has two corresponding matching blocks, $I_{n-1}(x+x_{nb},y+y_{nb})$ and $I_{n+1}(x+x_{nf},y+y_{nf})$ in the other two frames. Where (x_{nb}, y_{nb}) and (x_{nf}, y_{nf}) are the motion vectors associated with block $I_n(x,y)$ in frames, I_{n-1} and I_{n+1} respectively. Here, x and y corresponds, to block indices in horizontal and vertical directions, respectively. Top module of Figure 4.4 explains the details of the initial motion block reconstruction.

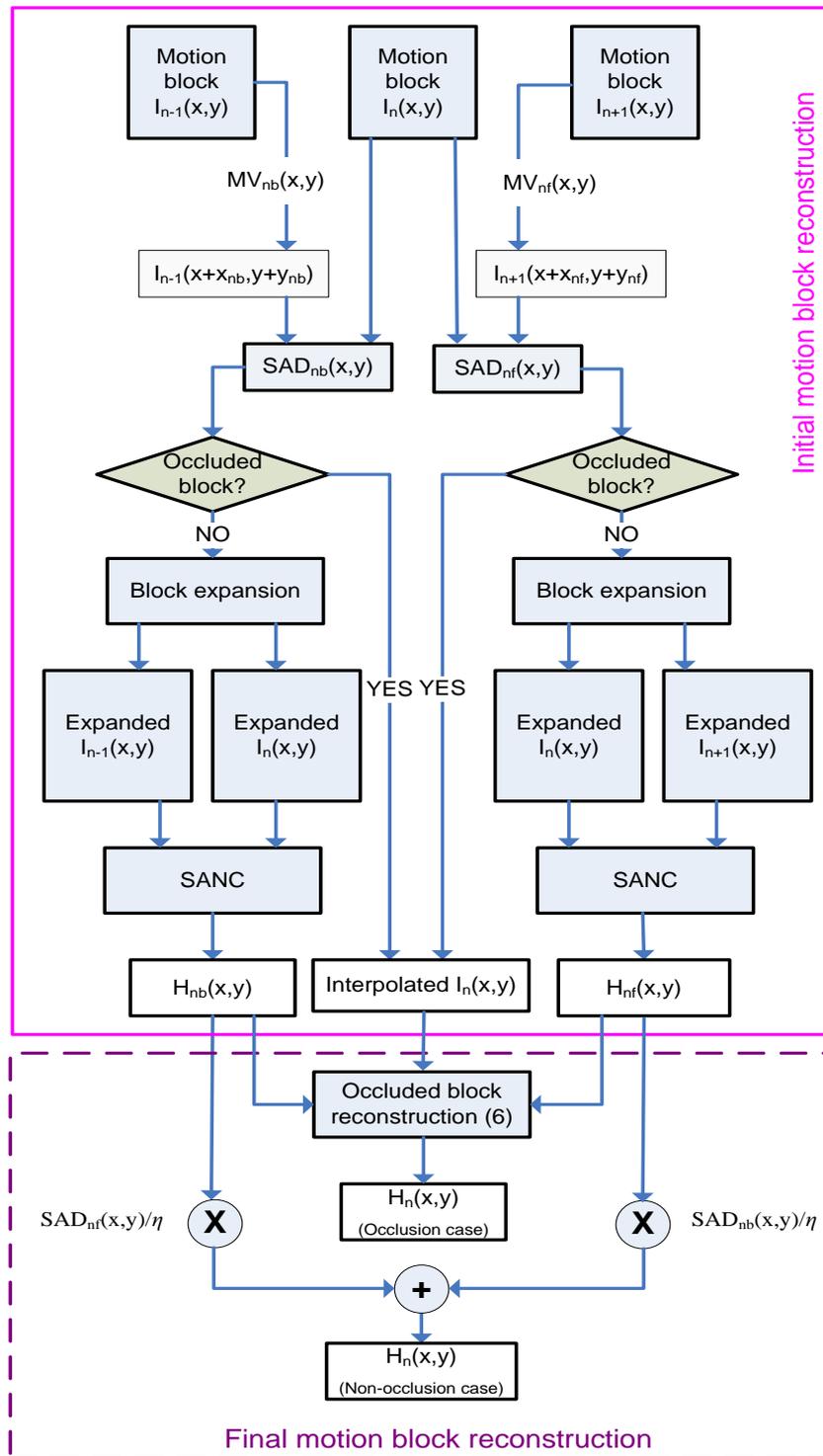


Figure 4. 4: Block diagram of the motion block processing consisting of initial and final motion block reconstruction.

In each frame there exist some blocks with no similar blocks in the neighboring frames. These blocks are declared as occluded blocks. Occlusion refers to the appearance of new objects and disappearance of the existing objects when comparing frames. Super resolving the occluded blocks is a discussion which is not considered in many super resolution algorithms. However occluded blocks can be found in many frames of the video sequences. Detecting and super resolving these blocks with a proper approach, results in a frame with higher resolution.

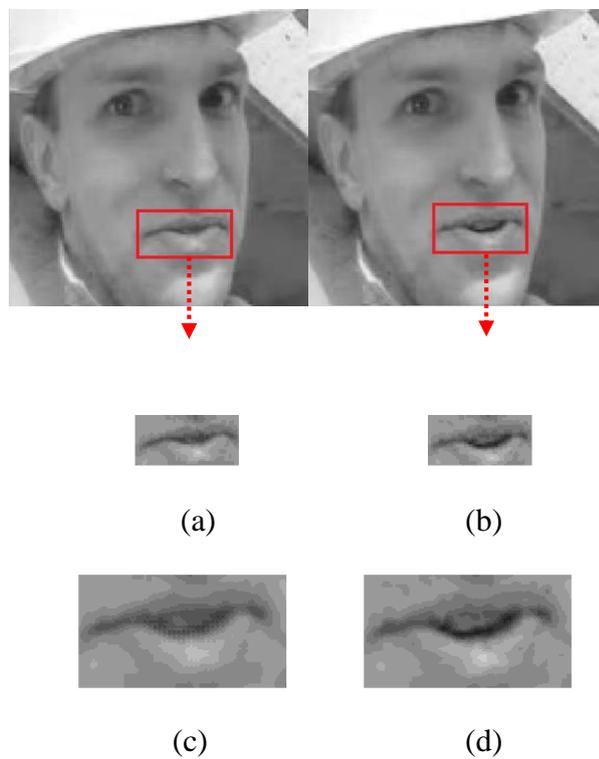


Figure 4. 5: Occluded area in frame 90 and 91 of Foreman video sequence. (a) and(b) show occluded areas in frames 90 and 91 respectively. (c) Super resolved area using Keren et al.[3] registration and SANC reconstruction[55]. (d) Interpolated area using DWT based interpolation [72].

Figure 4.5 shows frame numbers 90 and 91 of the Foreman video sequence. In frame 90 and 91 the corresponding area is regarded as occluded due to the fact that in frame 90 there is no teeth where in frame 91 teeth are noticeable. The appearance of the new object, which in this case is teeth, causes occlusion. The proposed technique employs DWT based interpolation in order to obtain the high resolution occluded areas. figure 4.5 (c) and (d) show the super resolved occluded blocks in figure 4.5 (a) and (b) using a classical SR (Keren et al.[3] registration and SANC reconstruction [55])and DWT based interpolation [72] respectively. Given that frame 91 is the reference frame, as you can see interpolation based method outperforms the super resolution in the case of occluded areas.

In order to identify the occluded blocks, the motion compensated error (MC error) is calculated for each block using the sum of absolute differences (SAD). The formula is as follows:

$$SAD_{nb}(x, y) = \sum |I_n(x, y) - I_{n-1}(x + x_{nb}, y + y_{nb})| \quad (27)$$

$$SAD_{nf}(x, y) = \sum |I_n(x, y) - I_{n+1}(x + x_{nf}, y + y_{nf})|$$

The computed MC errors are compared with an adaptive threshold $T_n(x,y)$, so that if the compared error is larger than the threshold the block is declared to be occluded.

if $(SAD_{nb}(x, y) \geq T_n(x, y)) \& (SAD_{nf}(x, y) \geq T_n(x, y)) \Rightarrow$ occluded

Typically, the MC error is proportional to the variance, $\sigma_n(x,y)$, of the absolute difference (AD) of $AD_{nb}(x,y)$ and $AD_{nf}(x,y)$. A threshold is adaptively determined as follows:

$$T_n(x,y) = \alpha \cdot \sigma_n(x,y) + T_m \quad (28)$$

T_m and α , are heuristically founded to be 140 and 0.8 using several training frames which are different than the test video sequences. T_m is the average of all the thresholds that the proposed technique obtains the best PSNR for the training frames. α is determined from the boundary between the lower 90% and upper 10% of the histogram of σ .

After separating the occluded and non-occluded motion blocks, a state of the art reconstruction method is used to enhance the resolution of the non-occluded motion blocks and form the high resolution image H_{nb} and H_{nf} . These reconstructed blocks will be inserted into the super resolved background. However, a main drawback of reconstruction algorithms is the generated artifacts within a maximum 4 pixels distance of the boundary of reconstructed image. To overcome this problem, the size of the blocks is expanded to $2+\lambda+2$ by $2+\lambda+2$ which after the reconstruction process the central $\lambda \times \lambda$ pixels region of it is inserted into the final super resolved image. Where, λ is the enlarging factor. The details of these block expansions are shown in figure 4.6.

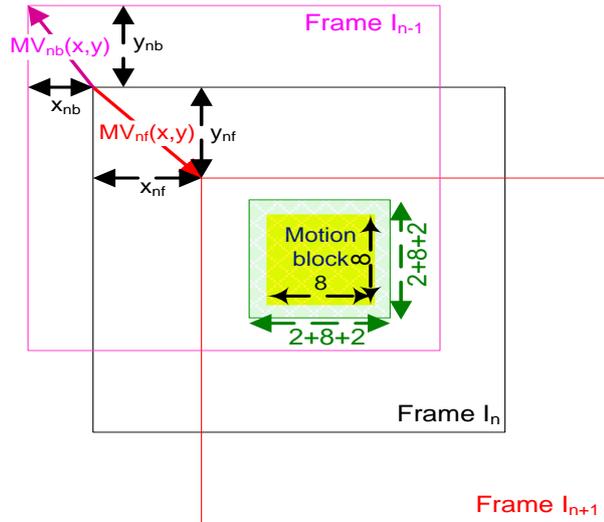


Figure 4. 6: Three consecutive frames I_{n-1} , I_n and I_{n+1} are aligned on a common motion block. The blocks are expanded by 2 pixels from sides in each frame.

Four motion blocks of the 88th frame of News sequence are shown in figure 4.7(a). Figure 4.7(b) is the reconstructed motion blocks without block expansion by the enlargement factor of 2. The observed artifacts around the boundary of the reconstructed block are disappeared in figure 4.7(c). This is achieved by applying block expansion before reconstruction. The blocks shown in figure 4.7 (c), are extracted from the central 16x16 pixels of the reconstructed expanded motion block.

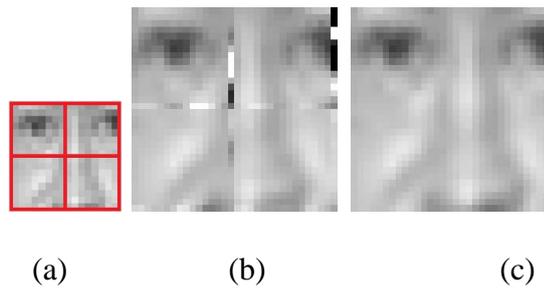


Figure 4. 7: A part of the 88th frame of News with four motion blocks: (a) reference blocks (b) reconstructed blocks without block expansion (c) reconstructed blocks with block expansion.

4.2.2.2 Final motion block reconstruction

After reconstructing the motion blocks in $H_{nb}(x,y)$ and $H_{nf}(x,y)$, it is time to combine them into a single high resolution motion block. $I_{n+1}(x+x_{nf}, y+y_{nf})$ and $I_{n-1}(x+x_{nb}, y+y_{nb})$ both contain information which help in reconstructing the high resolution block $H_n(x,y)$ from $I_n(x,y)$. In the proposed technique during the SR process the attempt is to involve more relevant information which can generate a block with higher resolution. It is clear that $I_{n+1}(x+x_{nf}, y+y_{nf})$ and $I_{n-1}(x+x_{nb}, y+y_{nb})$ does not hold the same information. It is beneficial to involve the block with more relevant information to the reference frame, in the SR process. Bottom module of Figure 4.4 illustrates the details of the final motion block reconstruction.

In the proposed technique a larger weight is assigned to the block with more relevant information to the reference block. Determining the weights is of crucial importance that affects the quality of the output block. In this context, $SAD_{nb}(x,y)$ and $SAD_{nf}(x,y)$ are used to calculate the weight of resolution enhanced motion blocks (i.e. $H_{nf}(x,y)$ and $H_{nb}(x,y)$). Since the block with less SAD contains more relevant information to the reference block, a higher weight is assigned to the matching block and a formula is designated as follows:

$$H_n(x, y) = \frac{SAD_{nb}(x, y)}{\eta} H_{nf}(x, y) + \frac{SAD_{nf}(x, y)}{\eta} H_{nb}(x, y) \quad (29)$$

$$\eta = SAD_{nb}(x, y) + SAD_{nf}(x, y)$$

One of the problems in image registration is determining the motion vectors in the areas of occlusion. Many algorithms try to solve this problem [87-89] however; their main disadvantage is their high computational complexity. In spite of all these methods, rather than finding a motion vector for the occluded blocks, we try to reconstruct the block using the available blocks after interpolation. As mentioned interpolating an image produces more accurate result than super resolving it with wrong image registration. If block is located at an occluded area, the most efficient way is to interpolate the reference block. Recently wavelet based interpolation algorithms show outstanding results in comparison to other well-known interpolation methods. Wavelet-based methods [72, 76 and 90, 91] enhanced the image resolution by estimating the preserved high frequency information from the given images. They were based on the assumption that the image to be enhanced was the low frequency subband among wavelet-transformed subbands of the original one and the target is to estimate the high frequency subbands of wavelet transform, so that a resolution-enhanced image can be obtained. Using this fact, the occluded block is replaced by its interpolated LR block using DWT based interpolation [72].

The reconstruction of the occluded block or the block with wrong motion estimation is formulated as below:

$$H_n(x, y) = \begin{cases} H_{nf}(x, y), & \text{if } \begin{pmatrix} SAD_{nb}(x, y) \geq T_n(x, y) \\ \& \\ SAD_{nf}(x, y) < T_n(x, y) \end{pmatrix} \\ H_{nb}(x, y), & \text{if } \begin{pmatrix} SAD_{nb}(x, y) < T_n(x, y) \\ \& \\ SAD_{nf}(x, y) \geq T_n(x, y) \end{pmatrix} \\ H_{nd}(x, y), & \text{if } \begin{pmatrix} SAD_{nb}(x, y) \geq T_n(x, y) \\ \& \\ SAD_{nf}(x, y) \geq T_n(x, y) \end{pmatrix} \end{cases} \quad (30)$$

where, $H_{nd}(x, y)$ denote the high resolution interpolated block $I_n(x, y)$.

After, all of the motion blocks in the high resolution frame are generated; these blocks are inserted into the super resolved static background. This process completes the respective high resolution frame which will go through the post processing in the next module. Bottom part of figure 4.4 illustrates the block diagram of the final motion block reconstruction.

4.2.3 Post processing module

Both high resolution motion and static blocks are generated using same super resolution algorithm. However the occluded blocks are resolution enhanced using DWT based interpolation method. Therefore, at the boundaries between the interpolated occluded blocks and super resolved motion blocks, we can observe blocking artifacts, as the blocks are derived from different algorithms. To decrease the blocking artifacts, a smoothing filter is applied to the boundary pixels (Figure 4.8) [92].

This is done by averaging the pixels at the boundaries of the neighboring blocks. The remaining pixels at the center of the blocks stay unchanged.

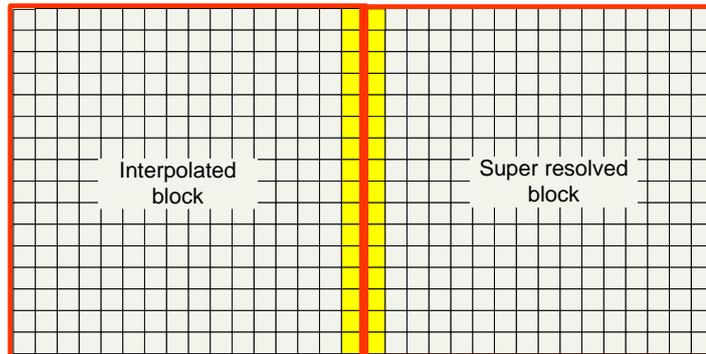


Figure 4. 8: The boundary (yellow) pixels of two neighbor blocks filtered using averaging.

The loss of high frequency components produces blurring. If the high frequency components could be restored, then image would posse its normal contrast for the relatively smaller objects. A sharpening process is applied to restore the high frequencies by adding a high frequency layer to the initial estimated high resolution frame H_n . As presented in [72] since the LR image is processed at different high frequency channels, the interpolated H_{nd} carries extra information at its high frequencies. Thus, in the following algorithm we add the high frequencies of the H_{nd} to H_n in order to improve the resolution of H_n . The steps of the algorithm are as follows:

Step 1: H_{nd} is decomposed into different low and high frequency subbands using discrete wavelet transform (DWT).

Step 2: At the LL subband the coarse features that are illustrated by low frequencies exist. On the other hand, fine details presented by higher frequency are available in the high frequency subbands. The information in the LL subband is not required. Thus, by placing zero instead of LL subband all the low frequency information will be removed.

Step 3: Composing all the subbands using inverse DWT (IDWT), generates a full high frequency image (HH_{nd}).

Step 4: Finally, enhancement of the high frequency details is achieved by adding the high frequency layer (HH_{nd}) to the initial estimated high resolution frame H_n . This produces an end result SR_n in which the high frequencies have been boosted closer to the original high resolution frame and the low frequencies remain as before.

Figure 4.9 shows the sharpening algorithm in details.

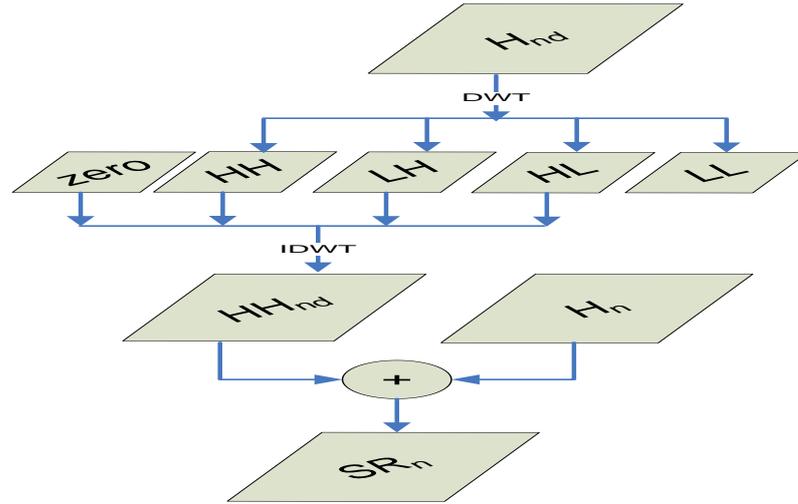


Figure 4. 9: Block diagram of the sharpening process.

4.3 Experimental Results

In order to show the performance of the proposed technique, four benchmark video sequences, Ice, Foreman, News and Container are used in this section. The sizes of the LR video sequences are 256x256 pixels which are the blurred and down-sampled version of the original videos of size 512x512 pixels. The frame rate of all the test video sequences is 30 frames per second and each of the videos has 200 frames.

At the pre-processing module the motion blocks are identified using the optical flow algorithm [34]. A motion vector is assigned to each block of equal size in the LR reference frame according to its displacement in the neighboring frames. First experiment is dedicated to determine an optimum size for the blocks. The density of motion fields is decreased from one motion vector per 8x8 pixel blocks to one motion vector per 16x16, 32x32 and 64x64 pixel blocks. The minimum block size is chosen as 8x8 because this is the limit size for a block to be reconstructed using the reconstruction method [55]. Each motion block and its matching block in the neighboring frame are found using the computed motion vector. The average PSNR of

each motion block and its matching blocks are calculated and shown in Table 4.1. As can be noticed, the highest PSNR is achieved for the blocks of size 8x8 pixels since large block sizes decreases the motion estimation accuracy. Therefore, the block size 8x8 pixels is chosen for identifying and reconstructing the motion blocks in the proposed SR technique.

Table 4. 1: Average PSNR of the motion compensated blocks with different sizes in the various test sequences.

BLOCK SIZE	AVERAGE PSNR VALUES FOR BENCHMARK VIDEO SEQUENCES (dB)			
	FOREMAN	NEWS	CONTAINER	ICE
64x64	18.80	27.44	28.24	20.02
32x32	21.62	30.81	31.73	22.45
16x16	25.70	32.48	33.51	25.02
8x8	29.10	34.45	35.34	29.82

In this experiment the results achieved by applying “final motion reconstruction” is discussed. As stated before, a weight is assigned to each of the reconstructed blocks ($H_{nb}(x,y)$ and $H_{nf}(x,y)$) for final reconstruction process. These weights are given to each block according to their similarity with the reference block. To show the effect of the assigned weights, the blocks are once reconstructed without any weight. This is done by simply averaging the blocks. The results are compared with the reconstructed weighted blocks. Table 4.2 illustrates the average PSNR of this comparison. The obtained PSNR values approve the superiority of the reconstructed weighted blocks over the averaged blocks.

Table 4. 2: Average PSNR of the reconstructed blocks with and without assigning weight.

Weight	AVERAGE PSNR VALUES FOR BENCHMARK VIDEO SEQUENCES (dB)			
	FOREMAN	NEWS	CONTAINER	ICE
without weight	39.35	37.18	36.22	38.65
with weight	39.90	37.43	36.46	38.85

As discussed before the high frequency components of the resolution enhanced input reference frame is added to the super resolved input LR frame. The average PSNR of the super resolved frames before and after high frequency addition is evaluated in the last two rows in Table 4.3. The comparison shows that DWT based interpolation contains useful information in which it increased the average PSNR of the video frames. This is due to the fact that DWT based interpolation has the ability to estimate the preserved high frequency information of the given LR frame with high accuracy.

We compared the performance of the proposed technique with those of six state of the art algorithms: NEDI [48], DWT based interpolation[34], Keren et al. registration[3] and SANC reconstruction[55], Marcel et al. registration[53] and SANC reconstruction and Vandewalle et al. registration[54] and SANC reconstruction. Emphasizing the edges (high frequency) of the images using DWT based interpolation [34], causes enhancement in the resolution of the images. This method preserves the edges of the interpolated image (using bicubic interpolation) in three orientations of 0, 45 and 90 degrees. By using directional high-pass filters, the performance of this method is evaluated and called as, HPF based interpolation in Table 4.3. To do this experiment, first, bicubic interpolation is applied to the low resolution frame. Afterwards, three high-pass filters with the directions similar to DWT is chosen, and the edges of the interpolated frame are emphasized in these directions. Table 4.3 shows

the average PSNR results of the super resolved test videos using different resolution enhancement methods. Note that the proposed technique provides better results in terms of PSNR measures and visual quality.

Table 4. 3: The average PSNR results of the super resolved test videos using different resolution enhancement methods (enlargement factor of 2).

Resolution Enhancement Method	AVERAGE PSNR VALUES FOR BENCHMARK VIDEO SEQUENCES (dB)			
	FOREMAN	NEWS	CONTAINER	ICE
NEDI[48]	35.01	32.29	31.41	34.68
Keren reg.[3]& SANC Recon.[55]	36.20	32.50	31.72	35.47
Marcel reg.[53]& SANC Recon.[55]	36.16	32.53	31.66	35.33
Vandewalle reg.[54]& SANC Recon.[55]	36.28	32.65	31.74	35.61
DWT [72]	37.02	34.54	33.72	36.80
HPF based interpolation	35.72	32.83	32.82	35.42
Proposed SR without sharpening	39.01	36.88	35.60	38.12
Proposed SR with sharpening	39.90	37.43	36.46	38.85

Figure 4.10 shows the frame by frame PSNR values of super resolved Ice sequence, using different algorithms. The superiority of the proposed technique over the conventional SR algorithms is due to the fact that the registration is performed more precisely, the occluded blocks are identified and treated properly and the loss of the high frequency is decreased.

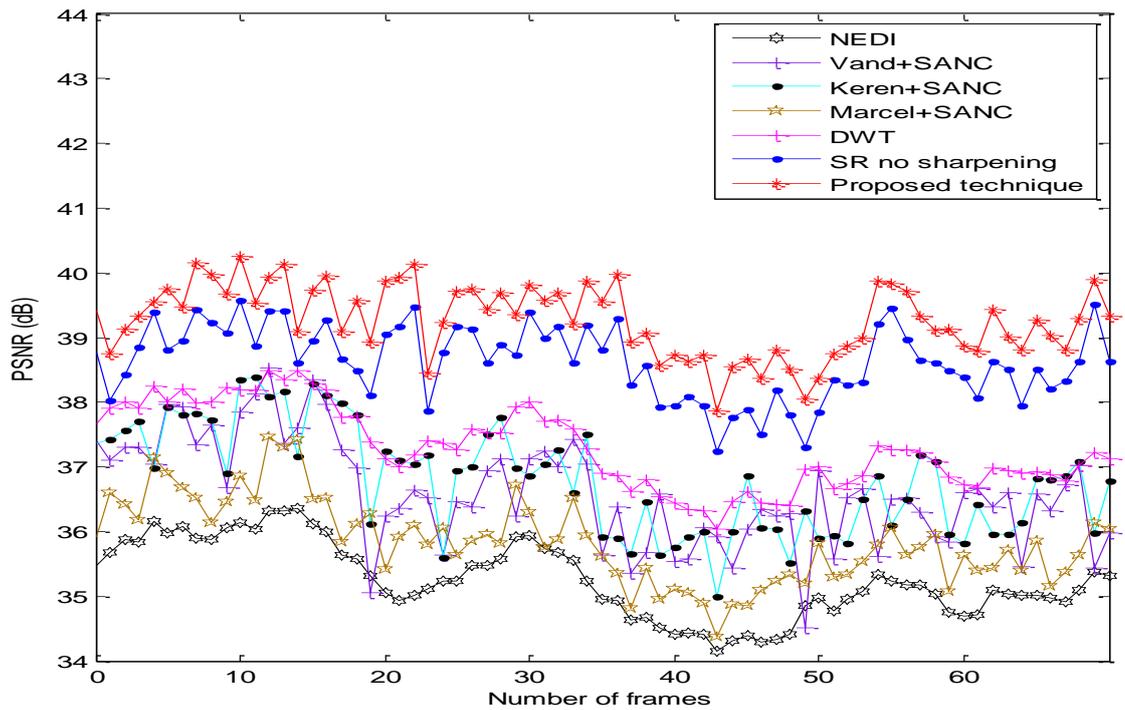


Figure 4. 10: Frame by frame PSNR results of resolution enhancement of “Ice” video sequence obtained from various resolution enhancement methods versus proposed technique.

Figure 4.11 and Figure 4.12 shows a part of super resolved 100th and 50th frames of Foreman and Container video sequences using various algorithms, respectively.

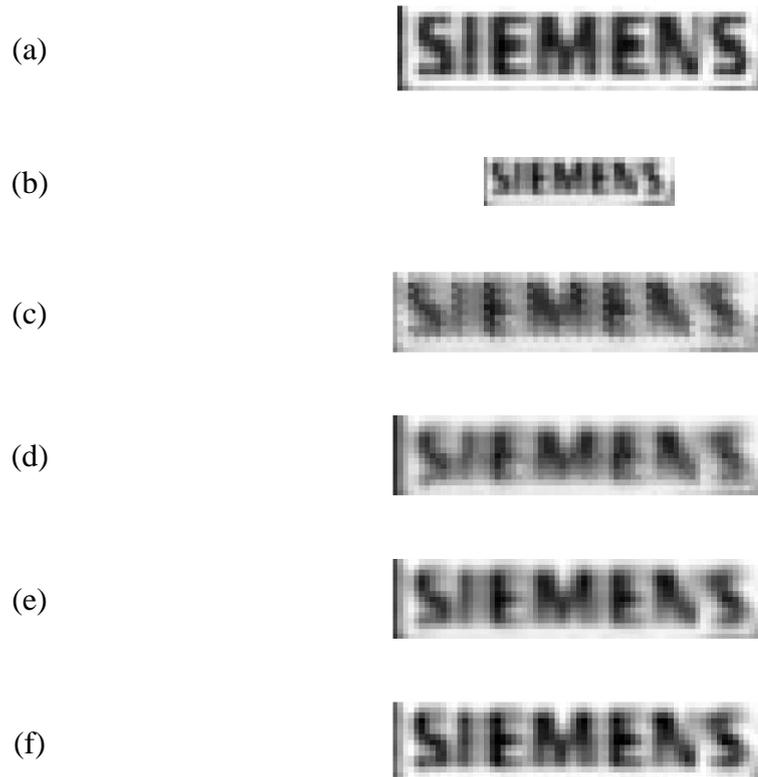
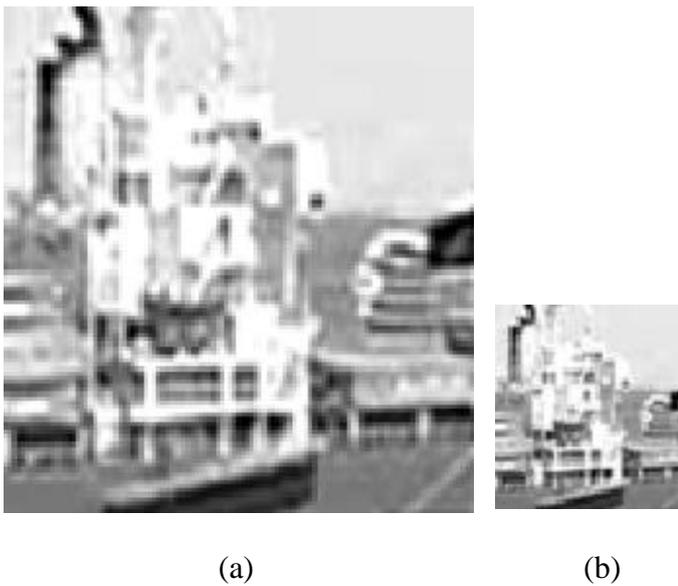


Figure 4. 11: A part of 100th frame of Foreman video sequence: (a) Reference HR image, (b) Input LR image, (c) Resolution-enhanced image by NEDI, (d) Marcel registration and SANC reconstruction, (e) DWT, (f) proposed technique.



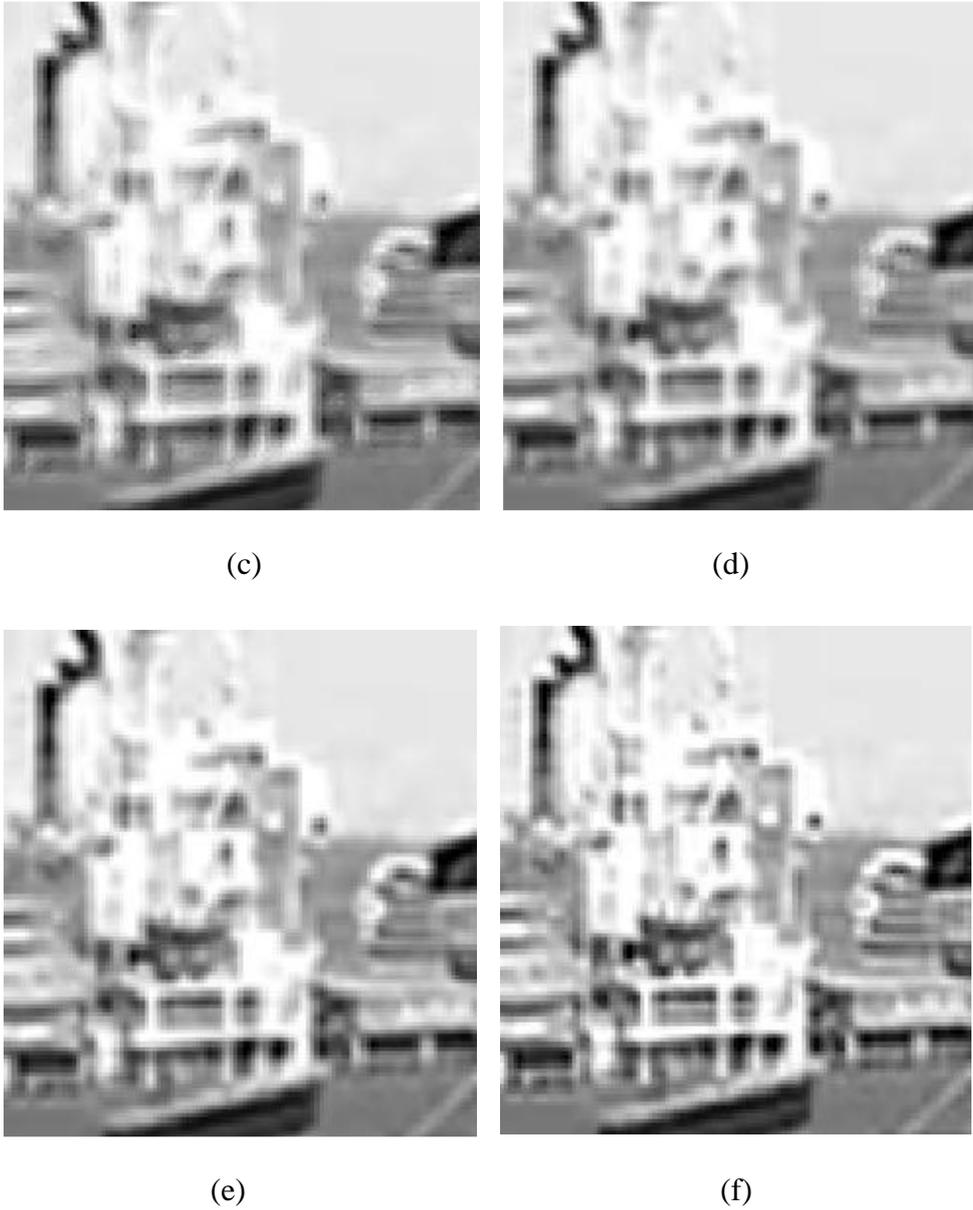
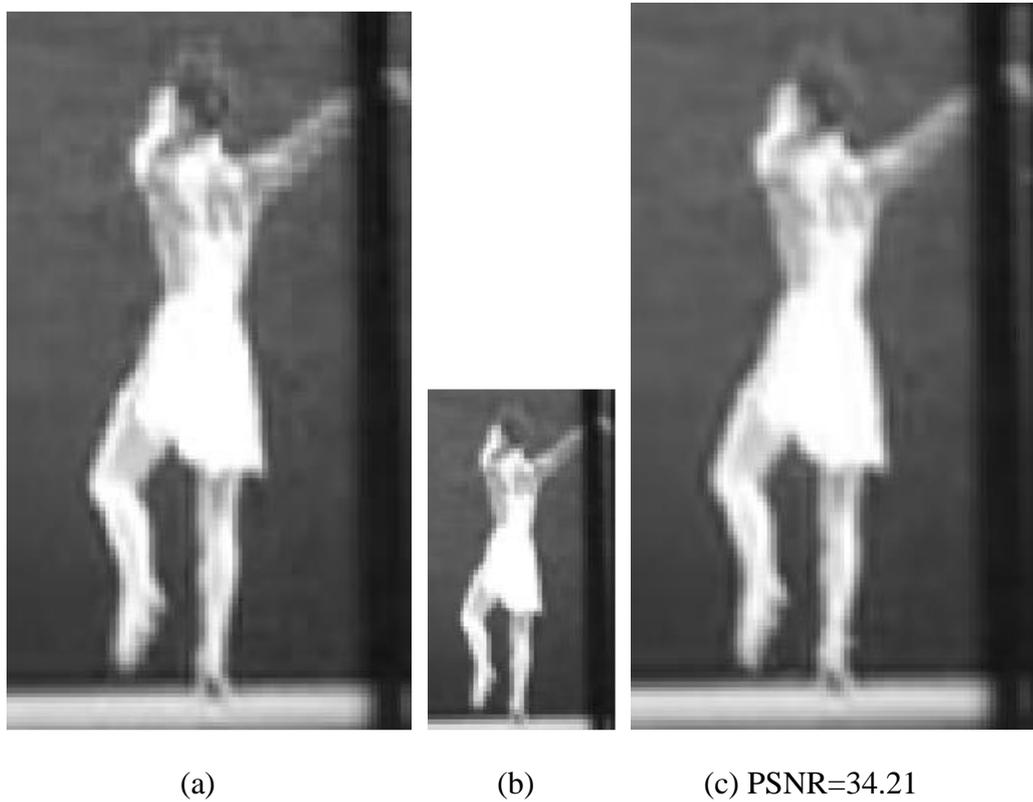


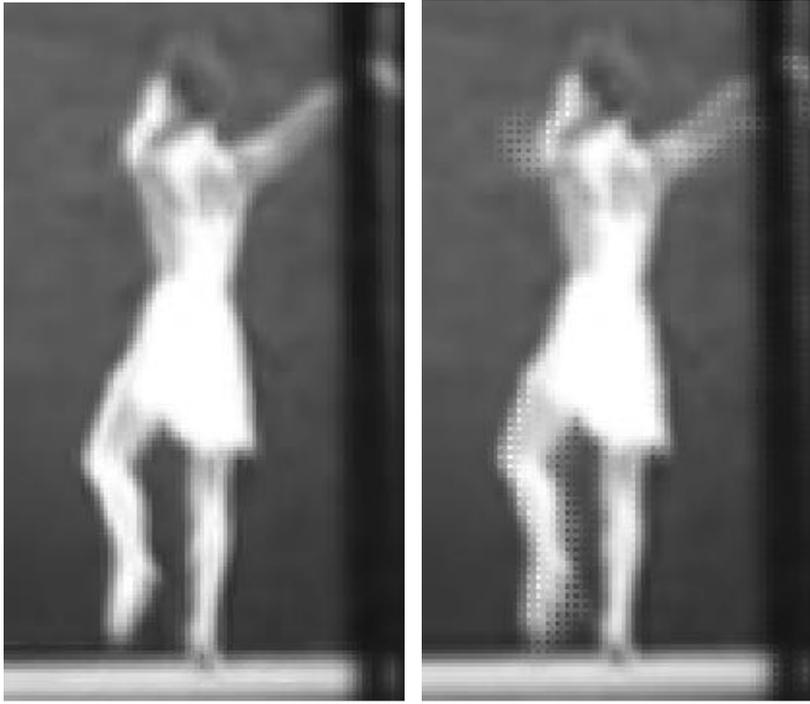
Figure 4. 12: A part of 50th frame of Container video sequence: (a) Reference HR image, (b) Input LR image, (c) Resolution-enhanced image by NEDI, (d) Keren reg. and SANC rec., (e) DWT, (f) proposed technique.

In order show the importance of accurate registration, different part of 3rd frame is illustrated in figure 4.13 and figure 4.14. Figure 4.13 contains motion blocks in which the erroneous motion displacements are estimated using Vandewalle registration method. The quality of the reconstructed part using NEDI interpolation is better than

the one reconstructed by Vandewalle-SANC super resolution. The PSNR of these parts are compared. As you can see, the proposed technique generates the best results; however because of inaccurate registration of Vandewalle registration method NEDI outperforms the Vandewalle-SANC super resolution. Figure 4.14 shows another part of the same frame in which the registration using Vandewalle registration method is more precise. As you notice, in this case Vandewalle-SANC super resolution outperforms NEDI in both PSNR value and visual quality.

Note that the proposed technique achieves much better visual and PSNR results in all the cases.





(d)PSNR=31.96 dB

(e) PSNR=24.65

Figure 4. 13: A part of 3rd frame of News video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by proposed technique (d) NEDI (e) Vandewalle registration and SANC reconstruction with incorrect registration.

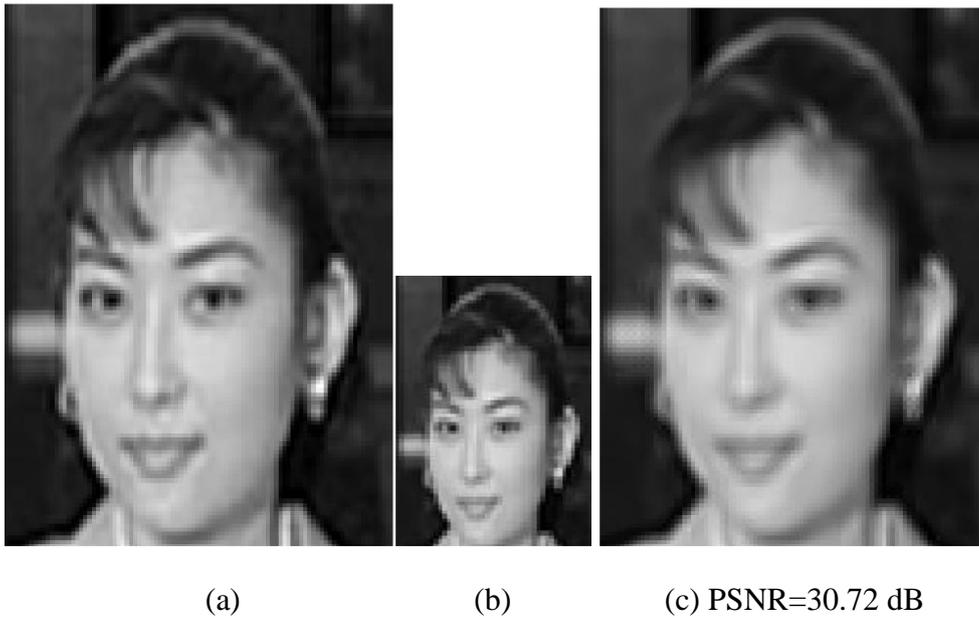


Figure 4. 14: A part of 3rd frame of News video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by NEDI (d) Vandewalle registration and SANC reconstruction with correct registration (e) proposed technique.

In chapter 3, we have demonstrated various super resolution techniques called MSR, MBSR and MBSR using NEDI, in which the size of super resolved frames were

four times larger than the size of input low resolution frames. However, this chapter presented a super resolution technique that changes the size of input low resolution frames by a factor of two. In order to compare the performances of these techniques, we need to have the same enlargement factor. To do this, the proposed SR is repeated two times on the test video sequences and the average PSNR and SSIM results are shown in Table 4.4.

Table 4. 4: The average PSNR and SSIM results of the super resolved test videos using MSR, MBSR and MBSR using NEDI and Proposed SR methods (e. f: 4).

Resolution Enhancement Method	AVERAGE PSNR AND SSIM VALUES FOR BENCHMARK VIDEO SEQUENCES (dB)					
	FOREMAN		NEWS		ICE	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
MSR	0.9280	31.10	0.8913	25.02	0.9501	31.33
MBSR	0.9399	31.80	0.9008	25.89	0.9600	32.12
MBSR using NEDI	0.9402	32.21	0.9010	26.15	0.9601	32.4
Proposed SR	0.9605	32.99	0.9350	27.47	0.970	33.00

Figures 4.15, 4.16 and 4.17 show a part of super resolved 25th , 5th and 13th frames of Foreman, Ice and News video sequences using MBSR using NEDI and Proposed SR algorithms, respectively.

The proposed SR has the best performance when the enlargement factor is two. However, Table 4.4 and Figures 4.15, 4.16 and 4.17 show that the proposed SR achieves better visual, PSNR and SSIM results also for the enlargement factor of four in comparison to the other SR methods.



(a)



(b)



(c) PSNR=31.44 SSIM=0.9325



(d) PSNR=32.21 SSIM=0.9538

Figure 4. 15: A part of 25th frame of foreman video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by MBSR using NEDI (d) proposed technique.



(a)



(b)



(c) PSNR=30.02 SSIM=0.9399



(d) PSNR=31.02 SSIM=0.9554

Figure 4. 16: A part of 5th frame of foreman video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by MBSR using NEDI (d) proposed technique.



(a)



(b)



(c) PSNR=26.16 SSIM=0.9001



(d) PSNR=27.48 SSIM=0.9344

Figure 4. 17: A part of 13th frame of foreman video sequence, (a) reference HR image. (b) input LR image (c) Resolution-enhanced input LR image by MBSR using NEDI (d) proposed technique.

CHAPTER 5

CONCLUSION and FUTURE WORK

5.1 Conclusions

In this thesis a motion block based super resolution technique for low resolution video enhancement is proposed. Super resolution can be used for enhancing the resolution of images or video sequences. However, because of the error rate, super resolving video sequences is not a common procedure to enhance the resolution of video sequence. In order to overcome this problem we proposed a method by the help of multi frame super resolution approaches which performs less error.

The proposed technique presented a novel video super resolution technique based on the motion, occluded and static blocks of the input LR frames. Precise registration was obtained by dividing the frames into motion and static blocks using optical flow algorithm. The new SR technique proposed a new threshold calculation for identifying the occluded blocks based on the mean and the variance of the motion blocks. It solves the problem of occluded block reconstruction by replacing them with the interpolated blocks using DWT interpolation. A new sharpening process formed the super resolved output frame using the high frequency sub-bands of the DWT based interpolated reference frame.

The method implemented and investigated with a large number of test sequences. Experimental results demonstrated that, the proposed SR technique outperforms the conventional super resolution algorithms in terms of PSNR measures and visual quality. This is mainly due to the fact that the images are registered more accurately,

when compared with the other registration methods. Furthermore, it provides information about occlusion.

5.2 Future Work

Although the system that we have established provides better results in terms of PSNR and visual quality in comparison to the other methods, it is considered to have moderate computational complexity, which prevents it from being used in real-time applications. Using other programming languages such as C or C++ can be helpful for possible real time applications.

In addition, reducing the size of blocks in super resolving images, allows the system to function faster. Lack of information in a small block prevents a precise reconstruction in all investigated super resolution methods. We believe that applying a super resolution algorithm, which can work on pixels or use only the 2d motion object instead of the rectangular region not only can reduce the computational complexity but also can increase the quality of the super resolved image.

REFERENCES

- [1] P. Milanfar, “super resolution imaging “, *CRC press*, 490p, 2010.
- [2] R. Y. Tsai and T. S. Huang, “Multiframe image restoration and registration”,
in Advances in Computer Vision and Image Processing, T. S. Huang, Ed. JAI
Press, Vol. 1, pp. 317–339, 1984.
- [3] D. Keren, S. Peleg and R. Brada, “Image sequence enhancement using subpixel
displacements”, *IEEE Computer Society Conference on Computer Vision and
Pattern Recognition*, June 1988, pp. 742-746.
- [4] M. Irani and S. Peleg, “Improving resolution by image registration”, *CVGIP:
Graphical Models and Image Processing*, Vol. 53, No. 3, pp. 231–239, May
1991.
- [5] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, “High-resolution image
reconstruction from lower-resolution image sequences and space-varying
image restoration,” *Proceedings of IEEE International Conference on
Acoustics, Speech, and Signal Processing (ICASSP '92)*, vol. 3, pp. 169–172,
San Francisco, Calif, USA, March 1992.
- [6] M. K. Ng, J. Koo, and N. K. Bose, “Constrained total leastsquares
computations for high-resolution image reconstruction with multisensors,”
International Journal of Imaging Systems and Technology, vol. 12, no. 1, pp.
35–42, 2002.
- [7] M. K. Ng and N. K. Bose, “Analysis of displacement errors in high-resolution
image reconstruction with multisensors,” *IEEE Transactions on Circuits and
Systems, Part I*, vol. 49, no. 6, pp. 806–813, 2002.

- [8] M. K. Ng and N. K. Bose, "Fast color image restoration with multisensors," *International Journal of Imaging Systems and Technology*, vol. 12, no. 5, pp. 189–197, 2002.
- [9] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient super-resolution image reconstruction algorithm," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 573–583, 2001.
- [10] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Transactions on Image Processing*, vol. 3, no. 3, pp. 233–242, 1994.
- [11] L. Lucchese and G. M. Cortelazzo, "Noise-robust estimation of planar rotations with high precision", *Proc. IEEE Int. Conf. Image Process.*, vol. 2, Santa Barbara, CA, pp.699–7, Oct. 1997.
- [12] Michael Elad, Arie Feuer, "Super-resolution reconstruction of image sequences". *IEEE Trans. On Pattern Analysis and Machine Intelligence (PAMI)*, vol. 21, no. 9, September 1999.
- [13] Lin Z. and Shum H.Y., "On the fundamental limits of reconstruction-based super-resolution algorithms", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, , Vol. 1, pp. I-1171 - I-1176 CVPR 2001.
- [14] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.
- [15] Freeman W.T., Jones T.R., and Pasztor, E.C., "Example-based super-resolution", *IEEE Computer Graphics and Applications*, Vol. 22 Issue: 2, pp. 56 – 65, Apr 2002.

- [16] 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, pp. 1266–1271, August 1996.
- [17] 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, pp. 1769–1786, June 2000.
- [18] D. Capel and A. Zisserman, "Computer vision applied to super resolution", *IEEE Signal Processing Magazine*, Vol. 20, No. 3, May 2003, pp. 75–86.
- [19] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution", *IEEE Transactions on Image Processing*, Vol. 13, No. 10, pp. 1327–1344, October 2004.
- [20] I Daubechies, "Orthonormal Bases of Compactly Supported Wavelets", *Comm. Pure and Applied Math.*, 41, 909-996, 1988.
- [21] S G Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Trans. Pattern Reco. and Machine Int.*, 11(7), 674-693, 1989.
- [22] s.Grace Chang, Zoran Cvetkovic, and Martin Vetterli, "Resolution Enhancement of Images Using Wavelet Transform Extrema Interpolation", *IEEE ICASSP*, 2379-2382, May 1995.
- [23] Carey, W.K., Chuang, D.B., and Hemami, S.S., "Regularity preserving Image Interpolation", *IEEE Trans. Image Process.*, Vol 8, No.9, pp.1295-1297, 1999.
- [24] Kinebuchi, K., Muresan, D.D., and Parks, T.W., "Image Interpolation Using Wavelet-based Hidden Markov Trees". *Proc. ICASSP01*, Vol.3, pp.7-11, May 2001.

- [25] Woo, D.H., Eom, I.K., and Kim, Y.S., “Image Interpolation Based on Inter-scale Dependency in Wavelet Domain”. *IEEE Int. Conf. on Image Processing*, Oct. 2004.
- [26] M.S.Crouse, R.D.Nowak and R.G.Baraniuk, “Waveletbased Statistical Signal Processing Using Hidden Markov Models”, *IEEE Trans.on Signal Processing*, Vol. 46, No. 4, Apr.1998.
- [27] A.Temizel and T. Vlachos, “Wavelet Domain Image Resolution Enhancement using Cycle-spinning”, *Electronics Letters*, Vol.41, No.3, Feb. 2005.
- [28] A.Temizel and T. Vlachos, “Image Resolution Upscaling in the Wavelet Domain using Directional Cycle-spinning”, *Journal of Electronic Imaging*, Vol.14 No.4, Oct-Dec. 2005.
- [29] A.Temizel and T. Vlachos, “Wavelet Domain Image Resolution Enhancement”, *IEE Proc. Vis. Image Signal Processing*, Vol. 153, No.1, Feb. 2006.
- [30] Hasan Demirel, Sara Izadpanahi, and Gholamreza Anbarjafari,” Improved motion-based localized super resolution technique using discrete wavelet transform for low resolution video enhancement” *17th European Signal Processing Conference (EUSIPCO 2009)* , PP.1097-1101, Glasgow, Scotland, August 24-28, 2009.
- [31] S. Izadpanahi, H. Demirel,” Motion based video super resolution using edge directed interpolation and complex wavelet transform”, *Signal Processing, Elsevier*, Vol.93, Issue 7, pp.2076-2086, July 2013.

- [32] Huang, Shih-Chia, "An Advanced Motion Detection Algorithm with Video Quality Analysis for Video Surveillance Systems", *IEEE Trans. on circuit and systems for video technology*, vol. 21, No.1, Jan 2011.
- [33] S. Izadpanahi, H. Demirel, "Motion block based video super resolution", *Digital Signal Processing, Elsevier*, April 2013.
- [34] T. Brox, A. Bruhn, N. Papenberg and J. Weickert "High accuracy optical flow estimation based on a theory for warping", *In Proc. 8th European Conference on Computer Vision*, vol. 4, pp. 25-36, Prague, Czech Republic, May 2004.
- [35] Law Enforcement and Emergency Services Video Association. Retrieved 2010. <http://www.leva.org>.
- [36] Hongyan Zhang, Liangpei Zhang, Huanfeng Shen, "A super resolution algorithm for hyperspectral images", *Signal Processing, Elsevier*, Vol.92, Issue 9, pp.2082-2096, September 2012.
- [37] Ming-Hui Cheng, Hsuan-Ying Chen, Jin-Jang Leou, "Video super resolution reconstruction using a mobile search strategy and adaptive patch size", *Signal Processing, Elsevier*, Vol.91, Issue 5, pp.1284-1297, May 2011.
- [38] Hasan Demirel, Sara Izadpanahi, Golamreza Anbarjafari. "Video Resolution Enhancement by using Complex Wavelet Transform", *IEEE International Conference on Image Processing (ICIP-2011)*, Brussels, Belgium, Sep2011.
- [39] Hongyan Zhang, Liangpei Zhang, Huanfeng Shen, "A super resolution algorithm for hyperspectral images", *Signal Processing, Elsevier*, Vol.92, Issue 9, pp.2082-2096, September 2012.

- [40] Liangpei Zhang, Hongyan Zhang, Huanfeng Shen and Pingxiang Li, "A super resolution algorithm for surveillance images", *Signal Processing, Elsevier*, Vol.90, Issue 3, pp.848-859, March 2010.
- [41] S. P. Kim and W. Y. Su, "Recursive high-resolution reconstruction of blurred multiframe images", *IEEE Trans. Image Processing*, Vol. 2, pp. 534–539, 1993.
- [42] Michael Elad, Arie Feuer, "Super-resolution reconstruction of image sequences", *IEEE Trans. On Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 21, No. 9, September 1999.
- [43] R. Gonzalez and R. Woods, "Digital image processing", *Prentice Hall*, 2nd edition, 2002.
- [44] Resampling Methods. Retrieved 2010.
<http://seadas.gsfc.nasa.gov/help/general/ResamplingMethods.html>
- [45] Póth Miklós, Marka Oreškovića, "Image interpolation techniques" *2nd Serbian Hungarian. Joint Symposium (SISY 2004)*, October 2004.
- [46] J.D. van Ouwerkerk, "Image Super-resolution survey", *Image and Vision Computing*, vol. 24, pp.1039-1052, 2006.
- [47] Shaode Yu, Qingsong Zhu, Shibin Wu, Yaoqin Xie "Performance Evaluation of Edge-Directed Interpolation Methods for Images" *Computer Vision and Pattern Recognition*, March 2013.
- [48] X.Li and M.Orchard, "New edge-directed interpolation," *IEEE Trans. image process*, vol.10, no.10, pp.1521-1527, Oct.2001.
- [49] Edge-Directed Interpolation. Retrieved 2007,
<http://wodnrns.tistory.com/entry/Edge-Directed-Interpolation>

- [50] M. V. Joshi, S. Chaudhuri, and R. Panuganti, "Super-resolution imaging: use of zoom as a cue", *Image and Vision Computing*, Vol. 22, No. 14, pp. 1185–1196, December 2004.
- [51] Hasan Demirel, Sara Izadpanahi, "motion-based localized super resolution technique for low resolution video enhancement", European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP
- [52] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image by simultaneous registration, restoration and interpolation of low resolution images," in *IEEE Int. Conf. Image Processing*, Washington, DC, Oct. 1995.
- [53] 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.
- [54] P. Vandewalle, S. Süsstrunk and M. Vetterli, "A Frequency Domain Approach to Registration of Aliased Images with Application to Super-Resolution", *EURASIP Journal on Applied Signal Processing (special issue on Super-resolution)*, Article ID 71459, 2006.
- [55] T. Q. Pham, L. J. van Vliet, and K. Schutte, "Robust Fusion of Irregularly Sampled Data Using Adaptive Normalized Convolution", *EURASIP Journal on Applied Signal Processing*, Article ID 83268, 2006.
- [56] H. Knutsson and C.-F. Westin, "Normalized and differential convolution," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'93)*, pp. 515-523, New York, NY, USA, June 1993.

- [57] Joost van de Weijer, Lucas J. van Vliet, Piet W. Verbeek, Michael van Ginkel, "Curvature Estimation in Oriented Patterns Using Curvilinear Models Applied to Gradient Vector Fields", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.23 n.9, p.1035-1042, September 2001.
- [58] A. Patti, M. Sezan, and M. Tekalp, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. on Image Processing*, pp. 1064–1078, August 1997.
- [59] N. R. Shah and A. Zakhor, "Multiframe spatial resolution enhancement of color video". *Proc. of the IEEE Intl. Conference on Image Processing*, vol. I, Sep. 1996.
- [60] Hasan Demirel, Gholamreza Anbarjafari, Cagri Ozcinar, Sara Izadpanahi "Resolution Enhancement of Video Sequences by Using Discrete Wavelet Transform and Illumination Compensation", *Turkish Journal of Electrical Engineering and Computer Sciences*, August 2012
- [61] Sara Izadpanahi, Hasan Demirel, "Multi-frame super resolution using edge directed interpolation and complex wavelet transform", *IET Conference on Image Processing*, pp.1-5 July 2012.
- [62] G. Anbarjafari, S. Izadpanahi, and H. Demirel, "Video resolution enhancement by using discrete and stationary wavelet transforms with illumination compensation", *Journal of Signal, Image and Video Processing*, DOI: 10.1007/s11760-012-0422-1, (Accepted), Dec 2012.
- [63] C Sidney Burrus, R A Gopinath, and Haitao Guo, "Introduction to Wavelets and Wavelet Transforms: A Primer", Prentice Hall, NJ, 1998

- [64] S. Mallat, "A wavelet tour of signal processing", Published by Academic Press, 2nd edition ISBN 012466606X, 9780124666061, 1999.
- [65] Amara Graps, "Introduction to Wavelets", *IEEE Computational Sciences and Engineering*, 2(2), 50-61, 1995.
- [66] Panchamkumar D SHUKLA, "complex wavelet transform and their applications", Thesis of Master of Philosophy, 2003.
- [67] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution", *Proceedings on international conference on computer vision and pattern recognition (CVPR)*, pp. I-645 - I-650, 2001.
- [68] N G Kingsbury, "Complex Wavelets for Shift Invariant Analysis and Filtering of Signals", *Journal of Applied and Computational Harmonic Analysis*, 10, 3, 234-253, 2001.
- [69] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury. "The dual-tree complex wavelet transform". *IEEE Signal Process. Mag.*, 05(22):123–151, November 2005.
- [70] J. Yang, Y. Wang, W. Xu, and Q. Dai. "Image coding using dual-tree discrete wavelet transform". *IEEE Trans. Image Process.*, 2008.
- [71] T. Celik, H. Ozkaramanli, and H. Demirel. "Facial feature extraction using complex dual-tree wavelet transforms". *Comput. Vis. Image Understand.*, doi10.1016/j.cviu.2007.12.001, 2008.

- [72] Hasan Demirel and Gholamreza Anbarjafari, "Discrete Wavelet Transform Based Satellite Image Resolution Enhancement", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 1997-2004, Jun. 2011.
- [73] K. Jaya Priya, R.S Rajesh, "Image Resolution Enhancement using Inter-Subband Correlation in Wavelet Domain", *Special Issue for The International Conference on Computing, Communications and Information Technology Applications*, vol 5 no. 1, ISSN 1994-4608, Ubiquitous Computing and Communication Journal, *CCITA-2010*.
- [74] N.G. Kingsbury, "Image processing with complex wavelets", *Phil. Trans. Royal Society London A*, Vol. 357, pp. 2543-2560, Sept. 1999.
- [75] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern, C, Appl, Rev.*, vol. 34, no. 3, pp. 334–352, Aug. 2004.
- [76] H. Demirel and G. Anbarjafari, "Image Resolution Enhancement by using Discrete and stationary Wavelet Decomposition", *IEEE Trans. On Image Processing*, Vol. 20, no.5pp. 1458-1460, May 2011.
- [77] Pilla Jagadeesh and Jayanthi 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.
- [78] J.Allebach and P.W. Wong, "Edge-directed interpolation," *proceeding of ICIP*, pp.707-710, 1996.
- [79] L. D. Alvarez, J. Mateos, R. Molina, A. K. Katsaggelos, "High-resolution images from compressed low-resolution video: Motion estimation and observable

- pixels," *International Journal of Imaging Systems and Technology*, vol. 14, no. 2, pp. 58-66, 2004.
- [80] H. Shen, L. Zhang, B. Huang, 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.
- [81] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the non-local-means to super-resolution reconstruction," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 36-51, 2009.
- [82] Q.wang and R. K. Ward, "a new orientation adaptive interpolation method." *IEEE Trans. Image Process*, vol.16, no.4, pp.889-900, Apr.2007.
- [83] B.Lucas and T. Kanade. "An iterative image registration technique with an application to stereo vision". *In Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, Aug. 1981.
- [84] B. Horn and B. Schunck. "Determining optical flow". *Artificial Intelligence*, No.17. Pages 185–203, 1981.
- [85] H.-H. Nagel and W. Enkelmann. "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.
- [86] J. Weickert and C. Schnorr. "A theoretical framework for convex regularizers in PDE-based computation of image motion." *International Journal of Computer Vision*, 45(3):245–64, Dec. 2001.
- [87] S.-H. Lee, O. Kwon, and R.-H. Park, "Weighted-adaptive motion-compensated frame rate up-conversion," *IEEE Trans. Consumer Electronics*, vol. 49, no. 3, pp. 485–492, Aug. 2003.

- [88] B.-D. Choi, J.-W. Han, C.-S. Kim, and S.-J. Ko, "Frame rate up-conversion using perspective transform," *IEEE Trans. Consumer Electronics*, vol. 52, no. 3, pp. 975–982, Aug. 2006.
- [89] Y.-T. Yang, Y.-S. Tung, and J.-L. Wu, "Quality enhancement of frame rate up-converted video by adaptive frame skip and reliable motion extraction," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1700–1713, Dec. 2007.
- [90] Y. Piao, I. Shin, and H. W. Park, "Image Resolution Enhancement using Inter-Subband Correlation in Wavelet Domain", *International Conference on Image Processing*, Vol. 1, pp. I- 445-448, ICIP 2007.
- [91] Izadpanahi, Sara , Cagri Ozcinar, Gholamreza Anbarjafari, and Hasan Demirel. "DWT Based Resolution Enhancement of Video Sequences ", *Discrete Wavelet Transforms - A Compendium of New Approaches and Recent Applications*, edited by Awad Kh. Al - Asmari, ISBN 978-953-51-0940-2, InTech, February 2, 2013.
- [92] B. C. Song, S. C. Jeong and Y. Choi "Video super resolution algorithm using bidirectional overlapped block motion compensation and on the fly dictionary training" *IEEE Trans. On circuit and systems for video technology*, Vol. 21, No. 3, pp. 274-285, Mar 2011.
- [93] Fulya Erbay, Gozde B. Akar " A comparative evaluation of super resolution methods on colour images", Thesis of Master of Philosophy ,2011.
- [94] Yilca Baris Sert, Gozde B. Akar, Cagtay Candan" An examination of super resolution methods", Thesis of Master of Philosophy ,2006.

