Image Restoration Methods for Retinal Images: Denoising and Interpolation

by

Damber Thapa

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of philosophy
in
Vision Science

Waterloo, Ontario, Canada, 2015

AUTHOR'S DECLARATION

This thesis consists of material all of which I authored or co-authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

Retinal imaging provides an opportunity to detect pathological and natural age-related physiological changes in the interior of the eye. Diagnosis of retinal abnormality requires an image that is sharp, clear and free of noise and artifacts. However, to prevent tissue damage, retinal imaging instruments use low illumination radiation, hence, the signal-to-noise ratio (SNR) is reduced which means the total noise power is increased. Furthermore, noise is inherent in some imaging techniques. For example, in Optical Coherence Tomography (OCT) speckle noise is produced due to the coherence between the unwanted backscattered light. Improving OCT image quality by reducing speckle noise increases the accuracy of analyses and hence the diagnostic sensitivity. However, the challenge is to preserve image features while reducing speckle noise. There is a clear trade-off between image feature preservation and speckle noise reduction in OCT.

Averaging multiple OCT images taken from a unique position provides a high SNR image, but it drastically increases the scanning time. In this thesis, we develop a multi-frame image denoising method for Spectral Domain OCT (SD-OCT) images extracted from a very close locations of a SD-OCT volume. The proposed denoising method was tested using two dictionaries: nonlinear (NL) and KSVD-based adaptive dictionary. The NL dictionary was constructed by adding phases, polynomials, exponentials and boxcar functions to the conventional Discrete Cosine Transform (DCT) dictionary. The proposed denoising method denoises nearby frames of SD-OCT volume using a sparse representation method and combines them by selecting median intensity pixels from the denoised nearby frames. The result showed that both dictionaries reduced the speckle noise from the OCT images; however, the adaptive dictionary showed slightly better results at the cost of a higher computational complexity. The NL dictionary was also used for fundus and OCT image reconstruction. The performance of the NL dictionary was always better than that of other analytical-based dictionaries, such as DCT and Haar.

The adaptive dictionary involves a lengthy dictionary learning process, and therefore cannot be used in real situations. We dealt this problem by utilizing a low-rank approximation. In this approach SD-OCT frames were divided into a group of noisy matrices that consist of non-local similar patches. A noise-free patch matrix was obtained from a noisy patch matrix utilizing a low-rank approximation. The noise-free patches from nearby frames were averaged to enhance the denoising. The denoised image obtained from the proposed approach was better than those obtained by several state-of-the-art methods. The proposed approach was extended to jointly denoise and interpolate SD-OCT image. The results show that joint denoising and interpolation method outperforms several existing state-of-the-art denoising methods plus bicubic interpolation.

ACKNOWLEDGEMENTS

I would like to express my special thanks and appreciation to my supervisors, Dr. Kaamran Raahemifar and Dr. Vasudevan Laxminarayanan. Dr. Raahemifar has been very instrumental in giving direction to my research work and guiding me throughout the work. He is a true mentor and takes keen interest in his students. He is readily available to help anytime. His valuable inputs, guidance, criticism, sense of urgency, and his vast knowledge were paramount in completion of this work. He has supported me in every best way possible. Dr. Lakshminarayanan, has also provided me with the much needed insight for my research. His vast knowledge and experience has immensely enriched this work. His continuous encouragement helped me to foster new prospects and dimensions into my research work. It was an honor and a privilege to work under the joint supervision of Dr. Kaamran Raahemifar and Dr. Vasudevan Laxminarayanan.

I would also like to express my deepest gratitude to my committee members Dr. William Bobier and Dr. Simerjeet Saini, for their continuous support and guidance. Dr. Bobier also provided me with a comprehensive research support and grant in the first 2 years of my PhD program.

I would like to thank Mathiruban Tharmalingam for working with me at Dr. Raahemifar's lab and for sharing ideas with me. A very special thank you to my friend Abbas Ommani for sharing his clinical knowledge with me throughout the years. Thanks to all the faculties and staff of School of Optometry and Vision Science, University of Waterloo, especially Jennifer Casentino, Krista Parsons, Lisa Boxter, Anne Weber and Mirka Curran. Thank you to all my fellow graduate students at the University of Waterloo for your continuous support, without which this research would not have been completed.

Lastly, I am grateful to my family for their constant support and encouragement and to all my friends for their much needed support during this journey.

DEDICATION

To my Family.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMD	Age Related Macular Degeneration
ASDS	Adaptive Sparse Domain Selection
CCD	Charged Couple Device
OCT	Optical Coherence Tomography
CNR	Contrast-to-Noise Ratio
CRF	Conditional Random Field
CS	Compressive Sensing
CSR	Centralized Sparse Representation
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DR	Diabetic Retinopathy
DWT	Discrete Wavelet Transform
EB	Example-based
EGI	Edge-guided Interpolation
ENL	Equivalent number of looks
FOCUSS	FOCal Underdetermined System Solver
GSM	Gaussian State Mixture
HR	High Resolution
IBP	Iterative Back Projection
ICA	Independent Component Analysis
KLT	Karhunen-Loeve Transform
LPG-PCA	Local Pixel Grouping-Principle Component Analysis
LR	Low Resolution
MAP	Maximum A Posteriori
MRF	Markov Random Field
MSE	Mean Square Error
MSR	Mean-to-standard deviation ratio
MWNNM	Multi-frame Weighted Nuclear Norm Minimization
NA	Numerical Aperture
NCSR	Nonlocally Centralized Sparse Representation
NEDI	New Edge-directed Interpolation
NL	Nonlinear
NNM	Nuclear Norm Minimization
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis

pdf	Probability density function
POCS	Projection Onto Convex Set
PoEdges	Product of Edgeperts
PR	Photoreceptors
PSNR	Peak signal-to-noise ratio
RIP	Restricted Isometry Property
RPE	Retinal pigment epithelium
SAR	Simultaneous auto regressive
SD-OCT	Spectral Domain Optical Coherence Tomography
SME	Sparse Maxing Estimators
SNR	Signal-to-noise ratio
SOMP	Simultaneous Orthogonal Matching Pursuit
SR	Super-resolution
SSIM	Structural Similarity
SURE	Steins Unbiased Risk Estimator
SVD	Singular Value Decomposition
TD-OCT	Time Domain Optical Coherence Tomography
TV	Total Variation
VQ	Vector Quantization
WNNM	Weighted Nuclear Norm Minimization
XCOR	Cross-correlation

Chapter 1

Introduction

Examination of the retina provides a means to detect pathological changes that are both ocular and systemic in origin. Incidence of retinal pathologies, such as diabetic retinopathy (DR), glaucoma and age related macular degeneration (AMD) are sharply increasing, and in the case of DR, reaching epidemic proportions in western countries [1]. These pathologies can be detected through retinal imaging. Furthermore, telemedicine allows such retinal images to be obtained from remote areas and thereby allows diagnosis and treatment when a specialist is not present [2]. A well-known retinal imaging method is fundus photography. A fundus is the inner lining of the eye (the retina) and includes the macula, fovea, optic disc and retinal blood vessels. A fundus image of the right eye of a healthy patient taken with a fundus camera is shown in Figure 1.1. The image clearly delineates the retina, the retinal vasculature, and the optic disc from which the optic nerve connects the retina. The optic disc is normally an orange-yellow color circular area, measures about 1.5mm in diameter. The retinal arteries and veins that nourish the retina with oxygen emerge from the optic disc. The arterioles are brighter and narrower than venules. The macula is located approximately in the center of the retina. It is slightly darker than its surrounding area and is responsible for sharp central vision. Many retinal disorders can be diagnosed and followed up by examination of the fundus images. The fundus imaging also screens for the presence of early stages of the pathology and provides an enhanced analysis in clinical follow-up. Microaneurysms, hemorrhages, exudates and cotton wool spots, drusen, and abnormal and fragile new blood vessels are all indicators of retinal disease.

Optical Coherence Tomography (OCT) is one of the most popular and well-established non-invasive methods for imaging the interior of the eye [3]. OCT is similar to ultrasound; however, instead of sound it utilizes low coherence light source to image the tissue. By taking cross-sectional images of the retina, OCT provides detailed images of the retinal tissues

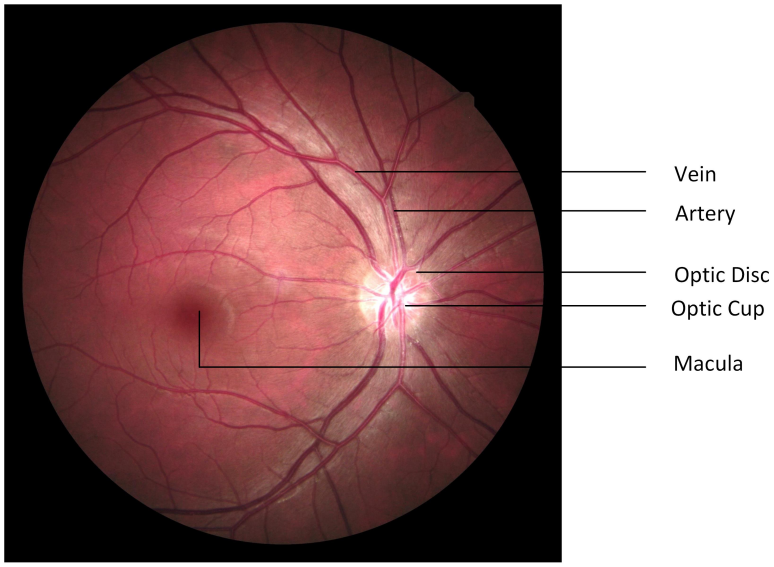


Figure 1.1: A fundus image of a visually normal eye.

and can capture subtle changes. The axial resolution of OCT ranges from 1-15 μm, which is 100 times higher than that of ultrasound. In addition, OCT provides cross-sectional images with imaging depth of 2-3 mm in tissue, and for that reason it became the most commonly used technique for imaging the interior of the retina. The OCT scan is captured in pseudo-color imaging; the red color represents the high reflectivity, green represents intermediate and blue/black represents low reflectivity. The red color usually comes from the reflectance of the junctions of inner and outer segments of photoreceptors (PR) and retinal pigment epithelium (RPE). Similarly, the green and blue/black colors come from the reflectance of inner and outer plexiform layers and nuclear layers, respectively. Figure 1.2 shows OCT scan printouts in horizontal and vertical directions of a healthy individual.

1.1 Optical Coherence Tomography

OCT creates an image by measuring the time delay and the strength of reflected light from the sample (tissue). It is performed using essentially a Michelson interferometer. A schematic diagram of time domain OCT (TD-OCT) is depicted in Figure 1.3. Light from the low coherence light source, such as broad-bandwidth laser or superluminescent diode,

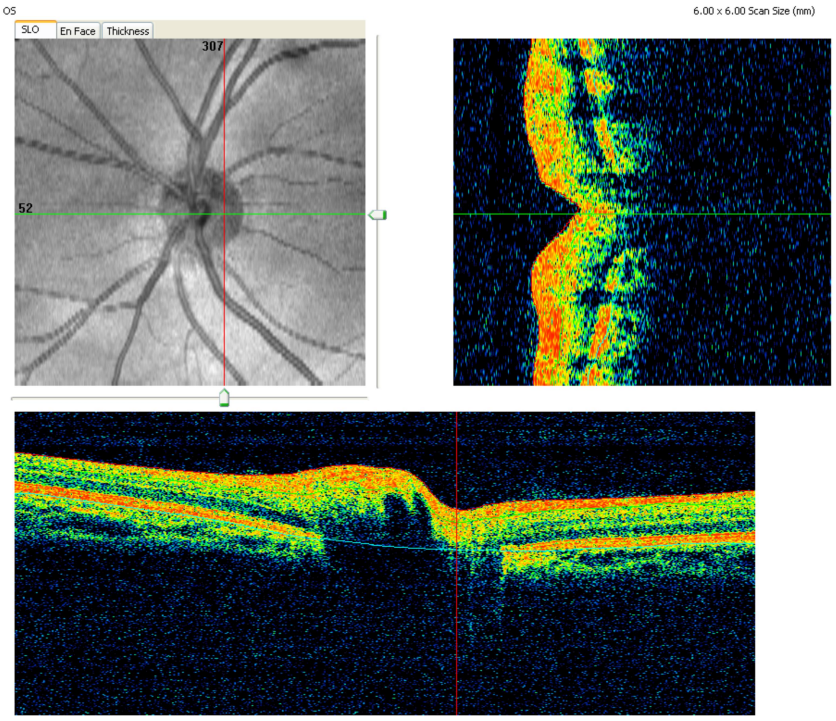


Figure 1.2: Optical coherence tomography image of optic disc taken from the left eye of a visually normal patient.

is divided into two arms by a beam splitter. One half of the light goes to a reference arm, which consists of a plane mirror at a known distance, and the other half goes to the sample arm of the interferometer, where it is reflected back by the tissue structures [4]. The light reflected back from the reference and sample arms are recombined by the beam splitter in order to produce an interference pattern. The constructive interference occurs only when the optical lengths of the reflected lights (from the mirror and tissue) are equal. Hence, the position of the reference arm is moved to allow imaging the tissue at different depths [5]. By scanning the reference mirror, a series of data from the different depths of the tissue along a point can be acquired which is known as an A-scan. A series of A-scans can be collected by shifting the position of the incident beam across the tissue. The collection of A-scans along the tissue creates a 2D cross-sectional image, which is known as a B-scan. Similarly, the collection of parallel B-scans produces a 3D volume. The image acquisition speed of TD-OCT depends upon the mechanical movement of the reference arm hence it is relatively slower than Spectral domain optical coherence tomography (SD-OCT).

SD-OCT employs a slightly different technique. A schematic diagram of SD-OCT is depicted in Figure 1.4. Instead of oscillating the reference mirror to capture the intensity of the reflected light at different depths, it keeps the reference arm stationary, and splits the interference pattern into its frequency components using a diffraction grating. The various depths in the tissue result in different frequency components in the interferogram. The intensities of the frequency components are simultaneously measured by the charged couple device (CCD). The SD-OCT simultaneously measures the spectral interferogram, hence the name spectral domain. The Fourier transform of the spectral measurement provides an A-scan, similar to that obtained from the TD-OCT. The SD-OCT does not require mechanical movement of the reference arm, rather it simultaneously collects all the frequencies, and therefore it is 100 times faster than the TD-OCT. In addition, it acquires a larger amount of data per second that results in a higher resolution. Therefore, SD-OCT has benefit over the TD-OCT both in terms of acquisition speed and resolution.

1.1.1 OCT Resolution

The term resolution refers to the ability of an imaging instrument to reveal the fine details in an object. In this thesis, the term resolution refers to the spatial resolution. Spatial resolution is further divided into axial and transverse resolution. Axial resolution is the ability to distinguish two closely-spaced points in the direction parallel to the direction of the incident beam (Figure 1.5). The axial resolution of OCT is determined by the

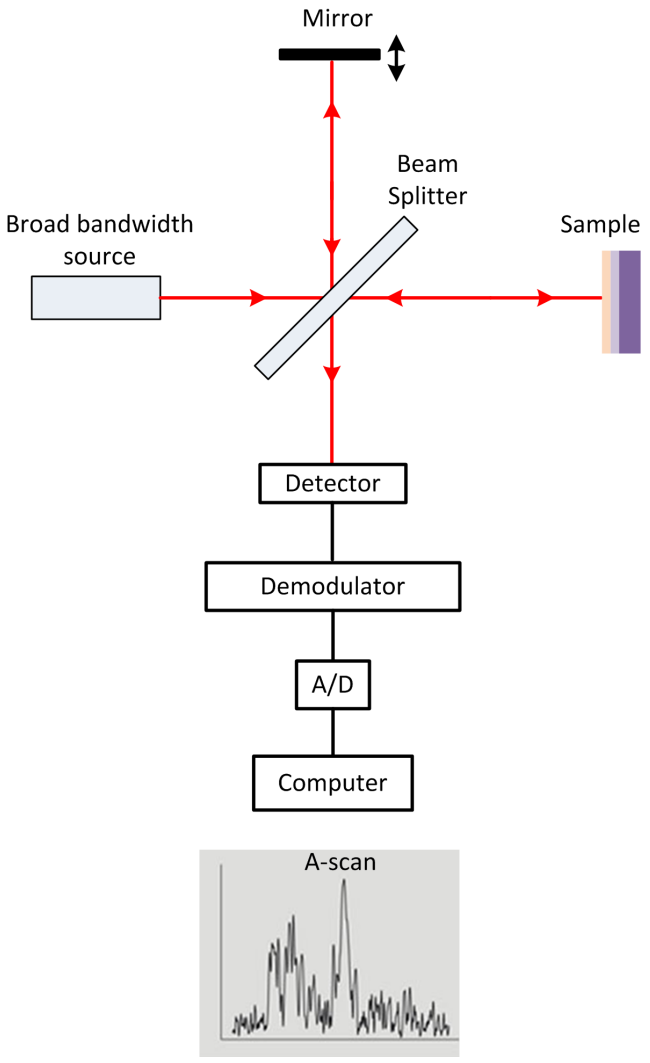


Figure 1.3: A schematic diagram of TD-OCT.

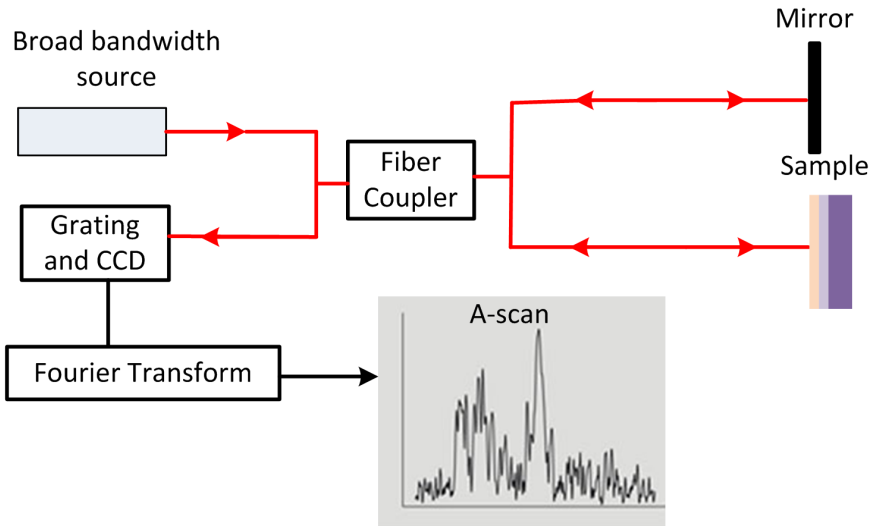


Figure 1.4: A schematic diagram of SD-OCT.

coherence length of the incident light and is given by

$$\Delta z = \frac{2\ln 2}{\pi}.\frac{\lambda^2}{\Delta\lambda} \tag{1.1}$$

where λ is the center wavelength of the input light and $\Delta\lambda$ is the spectral bandwidth. Since shorter coherence length occurs with the broader bandwidth light source, a high resolution (HR) image is obtained by using the broader bandwidth light. However, the size of the bandwidth is also limited, because the speed of the light in medium depends upon the wavelength, which causes extra delay and introduces additional phase difference if a very broad bandwidth is used. An HR image can also be obtained by using a shorter central wavelength light; however, the shorter wavelength light is scattered more by biological tissues. On the other hand, the transverse resolution is the ability of an instrument to distinguish two closely-spaced points in the direction perpendicular to the direction of the incident beam (Figure 1.5). The transverse resolution is determined by the size of the focused spot; the wider the focused spot, the poorer the transverse resolution. The transverse resolution is given by

$$\Delta x = \frac{4\lambda}{\pi}.\frac{f}{d} \tag{1.2}$$

where f is the focal length of the objective lens and d is the size of the incident beam. The transverse resolution of an imaging instrument depends mostly on its imaging optics, such

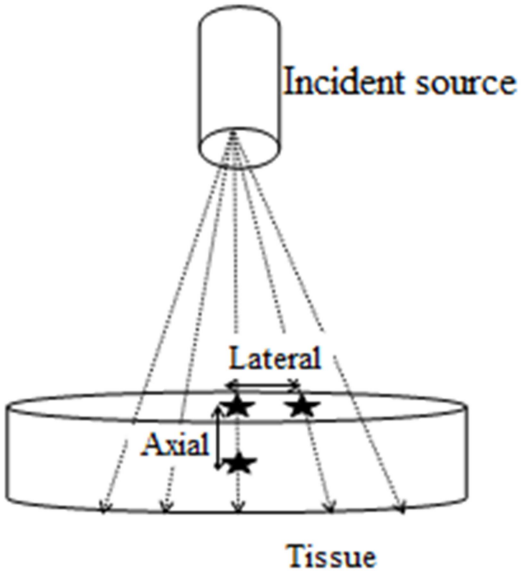


Figure 1.5: Illustration of axial and lateral resolution.

as numerical aperture (NA) [6]. The transverse resolution in terms of NA is given by

$$\Delta x = 0.37\frac{\lambda}{NA}. \tag{1.3}$$

A high NA lens can be used to generate a small spot size that can improve the transverse resolution; however, the resolution achieved is rather limited. Furthermore, both monochromatic and chromatic aberration increase with increasing NA [7]; therefore, it is not beneficial to increase the NA beyond a certain limit. In addition, there is a trade-off between the transverse resolution and depth of focus. The depth of focus (b) is given by

$$b = \frac{\pi(\Delta x)^2}{2\lambda}. \tag{1.4}$$

This relation shows that depth of focus decreases while increasing the transverse resolution, which limits the transverse resolution in the OCT. Figure 1.6 shows a schematic of sample arm optics of OCT.

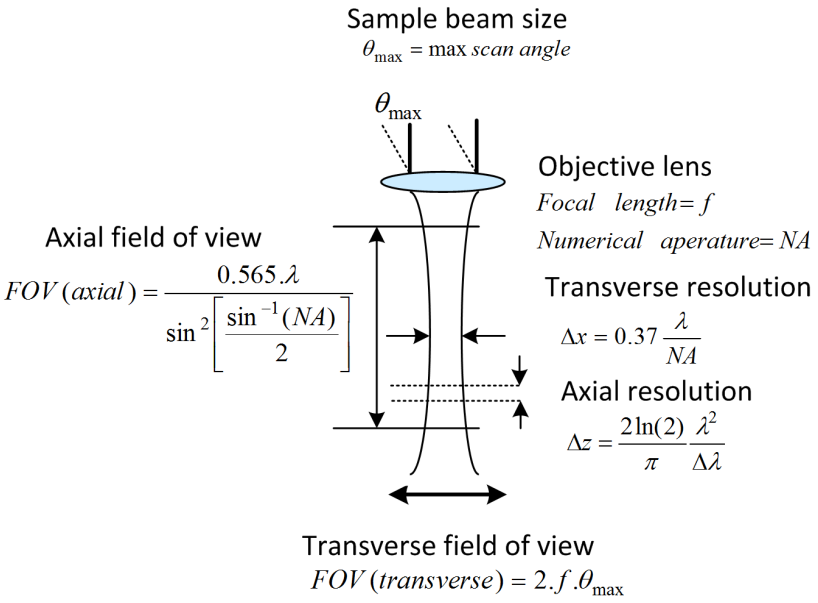


Figure 1.6: A schematic of sample arm optics of OCT (Adapted from [6]).

1.1.2 OCT Noise

The presence of noise degrades the quality of an image giving it a grainy, mottled, snowy or textured appearance. OCT noise is categorized into two types: system noise and speckle noise [7]. The system noise is the noise originating from any component of the imaging system, such as optical, electrical or optoelectrical components. Based on its origin, the system noise is further categorized into thermal noise, excess noise, shot noise and relative intensity noise [7]. The thermal noise, also referred to as the receiver noise, mostly arises from the energy exchange between the environment and the sensor. The excess noise is determined by the detector's quantum efficiency limit and originates from the multiplication process at a gain in the detector. The current fluctuation in the system produces shot noise. Shot noise increases when the input light power increases. The excess noise, spot noise and relative intensity noise vary when the powers of the light from the reference and sample arms of the OCT vary. The speckle noise arises from the coherence of unwanted back-scattered light. As mentioned earlier, OCT imagining modality relies on interferometry technique. It measures the spatial and temporal coherence of the back-scattering light from the reference and sample arms. However, this process also leads to speckle noise. Speckle noise is also results from an interference; however, it is produced by unwanted

back-scattered light and therefore has no obvious relation with the texture of the sample. Generally, the light reflected back from the tissue can be classified into three types: 1) the single-scattering light where the light reflected by a single particle is captured by the instrument; 2) the multiple-scattering light, where the light undergoes multiple scattering from more than two particles; and 3) the mixture of the unscattered and the multiple scattered light. Only the single-scattering light carries the information of the tissue; the latter two types lead to speckle noise. Many analytical models have been developed to eliminate the speckle noise from the signal. The most common model is based on the Beer-Lambert law that uses only the single-scattering light [5]. In this model the light intensity decays exponentially as

$$\hat{I}(z) = I_0 \exp(-2\mu_s z) \tag{1.5}$$

where μ_s is the scattering coefficient of a material, $\hat{I}(z)$ is the intensity of the back scattered light at the penetration depth z. $\hat{I}(z)$ can be assumed as the uncontaminated pixel value at depth z. Let $I(z)$ is the observed intensity (pixel) value at that depth. The observed intensity value can be obtained by modelling the speckle noise as the multiplicative random noise [8]

$$I(z) = \eta \times \hat{I}(z) \tag{1.6}$$

where η is a random variable whose value taken from the exponential probability density function (pdf) of the form

$$f_\eta(\eta) = \begin{cases} \exp(-\eta) & , \quad \eta \geq 0 \\ 0 & , \quad \text{otherwise.} \end{cases} \tag{1.7}$$

It is very challenging to differentiate the speckle noise from the speckles that carry microstructure information. The speckles that carry microstructure information originate from the large scatterers in the sample volume; whereas the speckle noise originates from the small scatterers outside the sample volume. The speckle noise can be reduced to some extent by suitably designing the optical scanner; however, it cannot be completely removed since the amplitude of speckle noise is the same as that of the signal carrying speckles [7].

1.2 Aims and Goals

The images taken with the OCT are contaminated by unwanted speckle noise. Since OCT imaging is based on interferometry, speckles are inherent in coherence imaging. The coherence between backscattered light produces speckle pattern in the image [9]. The speckle noise is visible as light and dark spots in the image. It reduces the contrast and

degrades the quality of the image, making the OCT images less informative in diagnostic examinations [9]. Therefore, to accurately detect the pathological changes, it is necessary to remove the speckle noise. However, the challenge is to retain image features while removing such noise. Besides removing the speckle noise, the resolution of the image should also be increased. An HR image has the ability to resolve micron scale retinal structures that allow a better understanding of the biophysics and visual process. An HR image contains more information and hence increases the accuracy in assessing the size, form and structure of a retinal lesion. An HR image requires a longer acquisition time at each lateral spot [10]. The longer acquisition time increases the motion artifacts due to the patients movement. Hence, a critical challenge in imaging by OCT is to optimally balance the trade-off between the image resolution, signal-to-noise-ratio (SNR), and the acquisition time. However, one may perform efficient post-processing techniques such as image super-resolution (SR) and denoising to reconstruct an HR high SNR image. Image interpolation and denoising are two widely used techniques in biomedical image processing. These techniques are mostly used one after another to reduce the computational speed. Although there exist several denoising and interpolation methods for OCT images in the literature, the search for efficient image denoising and interpolation methods is still an important and active research area. The main aim of this thesis is to develop new denoising and interpolation algorithms for OCT images which perform better than existing methods.

The raw images taken with fundus camera and OCT are very large. To store the images they are usually compressed. Image compression reduces the file's size. Telemedicine allows retinal images to be obtained from remote areas and thereby allows diagnosis and treatment when a specialist is not present. For online transformation of the medical images, the files need to be small, since smaller images can be transmitted faster and the increased transmission capacity reduces the transmission cost. To compress the image without sacrificing too much detail, it is necessary to use an appropriate dictionary (basis function). The second goal of this thesis is to develop a dictionary which yields better image reconstruction results for retinal images.

1.3 Contributions

Studies show that the performance of image reconstruction depends largely on the choice of the dictionary. Our first contribution is the development of a new analytical-based dictionary called nonlinear (NL) dictionary and its application for fundus and OCT image reconstruction. We also compare the NL dictionary with the other existing dictionaries for retinal image reconstruction. This work was done in collaboration with Mathiruban

Tharmalingam, a Masters degree student at the Ryerson University. The results of this study have been published in IEEE International Conference on Image Processing [11] and IEEE International Conference on Digital Image Processing [12].

Our second contribution is the design of a multi-frame image denoising algorithm using a sparse representation technique for reducing speckle noise from SD-OCT images. The algorithm denoises multiple B-scans of a SD-OCT volume using sparse representation technique and takes median pixels from the patches of the nearby B-scans to further suppress the speckle noise.

Our third contribution is the development of a joint speckle noise reduction and interpolation method for SD-OCT images using a low-rank approximation. This method uses weighted nuclear norm minimization (WNNM), and Singular Value Decomposition (SVD) fusion to recover an HR, high SNR image from a series of low resolution (LR), low SNR images extracted from a close location of SD-OCT volume. A partial results of this study have been accepted for publication in the Journal of Modern Optics [13].

SR is an off-line approach for improving the resolution of an image. Several methods have been developed for SR of natural images; however, these methods had never been applied to retinal images. We apply several SR methods on fundus images and compare their performances. The main results of this study have been published in Journal of Biomedical Optics [14].

1.4 Organization of the Thesis

The rest of the thesis is structured as follows. In Chapter 2, we present signal processing theories used in this thesis. We mostly concentrate on the sparse representation of a signal, compressive sensing and low-rank approximation. In Chapter 3, we present retinal image reconstruction using NL dictionary. In Chapter 4, we design a multi-frame sparsity-based SD-OCT image denoising method. In Chapter 5, we present a multi-frame speckle noise reduction method for SD-OCT images using low-rank approximation. Then, we design an image interpolation method for SD-OCT images using SVD fusion. We also present experiments which show how our denoising and interpolation method perform in comparison with the other existing techniques. In Chapter 6, we review several SR methods and demonstrate the positive impact expected from SR of retinal images and investigate the performance of various SR techniques. Chapter 7 concludes the thesis with some future directions.

Chapter 2

Theory

An image may be defined as a two-dimensional representation of an object. If the intensity values are measured on regular intervals along x and y directions and if the intensity values are finite and discrete, the image is called a digital image. Each element of a digital image is called a picture element or simply a pixel. Digital image processing methods perform some operations on digital images in order to enhance them or extract some useful information from them. These methods are applied to improve the pictorial information for human interpretation or to facilitate for autonomous machine perception. A common approach in digital image processing is to decompose an image into its fundamental components (such as Fourier transform) and then perform the required operations on each component separately or on a group of components. For example, Fourier analysis is an approach that decomposes an image into its sinusoidal components. The required processing is performed on Fourier coefficients. Thresholding is one of the most widely used operations for removing noise from an image. After performing the required operations, the image is reconstructed from its components. Table 2.1 and 2.2 show various types of transforms (parametric and non-parametric) [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31] that have been developed to decompose an image.

In this Chapter we overview some signal reconstruction techniques that are used for developing image restoration algorithms for this thesis. Section 2.1 presents signal reconstruction via sparse representation technique. Section 2.2 provides signal recovery techniques; such as compressive sensing and low-rank approximation, to reconstruct a signal/image from incomplete and corrupted measurements.

Table 2.1: A summarized table about atomic representations and sparse transforms: Parametric methods (non data-driven dictionary)

References	**Transform Domain**	**Dimensions**	**Parameters**	**Year**
Campbell and Foster [15]	Fourier Transform	1D/2D/3D	Frequency	1948
Ahmed, Natarajan and Rao [16]	Discrete Cosine Transform	1D/2D/3D	Duration translation	1974
Daubechies [17]	Separable Wavelet Transform	1D/2D/3D	Scale translation	1988
Coifman [18]	Wavelet Packet Transform	1D/2D/3D	Scale translation frequency	1992
Donoho [19]	Geometrical X-lets (wedgelet)	2D	Scale translation rotation	1997
Kingsbury [20]	Complex Wavelet Transform	2D/3D	Scale translation rotation	1998
Candes and Donoho [21]	Geometrical X-lets (curvelet)	2D/3D	Scale translation rotation	2000
Pennec and Mallat [22]	Geometrical X-lets (bandlet)	2D	Scale translation rotation	2004
Do and Vetterli [23]	Geometrical X-lets (contourlet)	2D	Scale translation rotation	2005
Lu and Do[24]	Geometrical X-lets (Surfacelet)	3D	Scale translation rotation	2005

Table 2.2: A summarized table about atomic representations and sparse transforms: Non-parametric methods (Dictionary learning)

References	**Transform Domain**	**Dimensions**	**Parameters**	**Year**
Jolliffe [25]	Principal Component Analysis	1D/2D/3D	Eigen Vectors	1986
Comon [26]	Independent Component Analysis	1D/2D/3D	Eigen Vectors	1994
Coifman and Lafon [27]	Diffusion Maps	1D/2D/3D	Spectral	2006
Engan and Aase [28]	Conventional Method of Optimal Directions	1D/2D/3D	Duration Translation	1999
Aharon [29]	K-SVD	1D/2D/3D	Duration Translation	2006
Coifman and Maggioni [30]	Diffusion Wavelet	1D/2D/3D	Scale Spectral	2006
Kafieh, Rabbani and Selesnick [31]	Complex Wavelet based K-SVD	1D/2D/3D	Scale translation rotation	2014

2.1 Sparse Representation of a Signal

Consider a finite dimensional discrete time signal $z \epsilon R^N$ and a basis $\Psi \epsilon R^{N \times N}$. The signal can be represented in terms of basis expansion with N coefficients $z = \sum_{i=1}^{N} \alpha_i \psi_i$, where ψ_i is the i^{th} basis vector and α_i is its coefficient. The effectiveness of the signal representation largely depends upon the choice of the basis function. Therefore, the basis is chosen in a way that the signal is represented by only $M << N$ basis vectors. Over the last century, there have been extensive studies on constructing a basis function that results in better signal representation. It has been found that the effectiveness of signal representation with the orthogonal basis is limited. An over-complete basis $D \epsilon R^{N \times K}; N < K$ (one with more columns than rows) has resulted in better signal representation [32]. A basis that has more columns than rows is called a *dictionary* and each basis function is called an *atom* [29]. A good dictionary can decompose a signal using a few atoms making signal representation

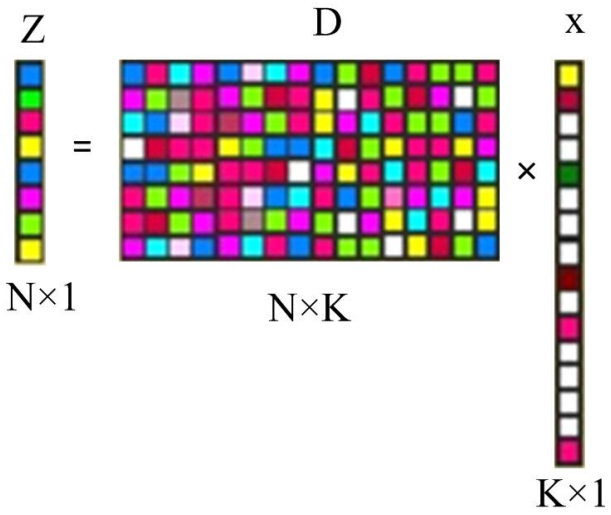


Figure 2.1: Sparse representation of a signal.

coefficients vector sparse, i.e., the coefficient vector has many zero coefficients and a few non-zero coefficients. This kind of signal representation is called *sparse representation* [29]. Sparse representation of a signal is obtained by solving the following optimization problem

$$\begin{aligned} \min_{x} \quad & \|x\|_0 \\ \text{subject to} \quad & Dx = z. \end{aligned} \tag{2.1}$$

The aim of sparse representation is to represent z using D such that both the number of non-zero coefficients $\|x\|_0$ and signal representation error $\|Dx - z\|_2^2$, where x is the sparse coefficient vector, are minimized. A schematic diagram of sparse coding is shown in Figure 2.1.

2.1.1 Sparse Coding Algorithms

Sparse representation searches for the most compact representation of a signal using a dictionary. This process involves solving equation 2.1 using an optimization algorithm. Equation 2.1 consists of l_0 norm which is equivalent to the number of non-zero coefficients in the coefficient vector x. The exact solution of this problem is an NP-hard because of its nature of combinational optimization; however, a suboptimal solution can be obtained

by iterative process using matching pursuits. An approximate solution can be acheived by converting l_0 norm to l_1 norm and solving the problem using basis pursuit. This subsection briefly discusses some pursuit algorithms.

Matching pursuit is a greedy algorithm that iteratively selects atoms from a given dictionary that have the highest inner product with the signal [33]. It consists of two stages: atom selection stage and the residual update stage. The atom selection stage finds the atom with the largest projection onto the signal (residual). In the residual update stage, the contribution due to that atom is subtracted from the residual to obtain a new residual. The process starts by setting the residual $r_0 = z$. If the residual at $(j-1)$ iteration is r_{j-1} then the residual in the j^{th} iteration is obtained by the following equation [34]

$$r_j = r_{j-1} - x_j d_j \tag{2.2}$$

where d_j is the atom selected in the j^{th} iteration and x_j is its corresponding coefficient. The process is repeated to the residual signal until the stopping criteria are met. The iteration stops if either the maximum number of atoms $\|x\|_0 < L$ is reached or signal representation error is smaller than the maximum error tolerance $\|Dx - z\|_2^2 < \varepsilon$. The matching pursuit is simple but it is slow and sometimes fails to converge due to sub-optimality. Orthogonal Matching Pursuit (OMP) [35] is a popular extension of matching pursuit in which the residual after each iteration is made orthogonal to all selected atoms by updating the coefficients, so no atom is selected twice [34]. Consider D_{j-1} is the matrix of atoms selected up to the $(j-1)^{th}$ iteration. Let d_j be the new atom selected in the j^{th} iteration such that the new matrix $D_j = [D_{j-1}, d_j]$ denotes the atoms selected up to the j^{th} iteration. The residual in the j^{th} iteration is given by

$$r_j = r_{j-1} - D_j(D_j^{\dagger} r_{j-1}) \tag{2.3}$$

where $D_j^{\dagger} r_{j-1}$ is the coefficient selected at the j^{th} iteration. Like matching pursuit the process is repeated to the residual signal until the stopping criteria are met. In OMP once an atom is selected, it will never be selected again in the successive iterations. Therefore, the algorithm converges fast. It should be noted that OMP requires more computation than matching pursuit; however, it leads to better signal representation. The residual update stage in the OMP involves pseudo-inverse transformation. This increases the computational complexity when the dictionary is large. To speed up the OMP, various algorithms have been proposed in recent years such as Cholesky OMP [36], Batch OMP [36] and Stagewise OMP [37].

Another well-known pursuit algorithm to solve the sparse representation problem is the Basis Pursuit [38]. Basis Pursuit finds the solution by minimizing l_1 norm of the coefficients

vector. The problem posed in equation 2.1 is reformulated by replacing the l_0-norm with l_1-norm as follows

$$\begin{aligned} \min_{x} \quad & \|x\|_1 \\ \text{subject to} \quad & Dx = z. \end{aligned} \tag{2.4}$$

The l_1 norm attempts to minimize the sum of the magnitude of the coefficients in the solution vector, i.e., $\|x\|_1 = \mid x_1 \mid + \mid x_2 \mid + \cdots + \mid x_N \mid$. The basis pursuit is the convex optimization problem and can be solved using linear programming methods. Basic pursuit has gained much interest in recent years especially in compressive sensing since it provides the sparest solution among all the possible solutions of an inverse problem. FOCal Underdetermined System Solver (FOCUSS) [39] is another sparse reconstruction algorithm similar to l_1-norm minimization; however, it gives the solution by minimizing l_p norm of the coefficient vector where $0 \leq p \leq 1$.

2.1.2 Dictionary Design

The signal representation error in sparse representation varies largely depending on the dictionary; therefore, the dictionary is chosen in a way that the error is minimized. The dictionary can be either chosen as a pre-specific set of functions (analytic-based dictionary) or learned from a training set to fit a given set of signals (learning-based dictionary) [29]. Construction of analytical-based dictionary is simple in which the atoms are created using a stationary function such as sine and cosine or wavelet functions. Time-Frequency Dictionaries, DCT, Gabor Transform, Wavelet Transform, Contourlet Transform, etc., have been used in the literature. Amongst these DCT and Wavelet Transform are commonly used. This subsection describes some of the dictionaries used in the sparse representation of a signal. In one dimension, DCT dictionary can be constructed using the following function

$$d_k(t) = \cos\Big[\frac{(2t+1)\pi k}{2N}\Big], k = 0, 1, \cdots, N-1. \tag{2.5}$$

The function creates a set of N atoms. The DCT dictionary can be extended to two dimensions by simply multiplying two 1-D atoms. Figure 2.2 shows an example of 2D DCT dictionary, with 256 atoms, that can be used to decompose a signal of length 64.

In 1990s DCT dictionary was widely used for many image processing applications, such as image compression (JPEG). However, the image reconstructed by DCT sometimes suffered from blocking effects as shown in Figure 2.3; therefore, it was later replaced by Discrete Wavelet Transform (DWT). The wavelets (wavelet basis functions) are created by

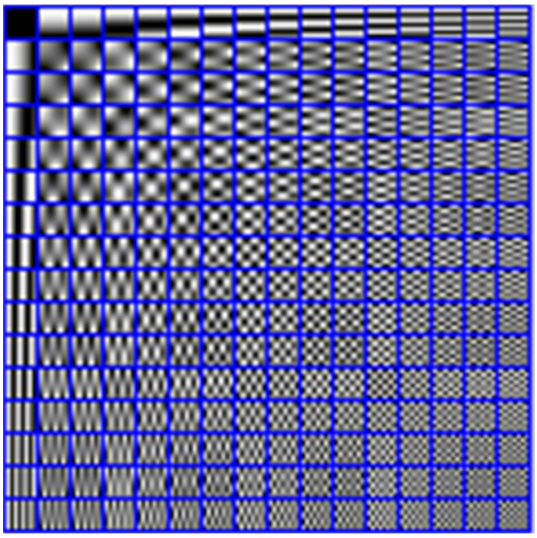


Figure 2.2: A DCT dictionary.

dilation and translation of the mother wavelet. For example, the following wavelet can be created by scaling mother wavelet by a factor of a and translating it by a factor of b

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\,\psi\big(\frac{t-b}{a}\big). \tag{2.6}$$

The simplest wavelet that can be created by scaling and translating mother wavelet is the Haar wavelet

$$\psi(t) = \begin{cases} 1 & 0 \leq \mathrm{t} \leq 1/2 \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

Figure 2.4 shows an example of 2D Haar dictionary, with 484 atoms, that can be used to decompose a signal of length 64. The wavelet offers a resizable structure for atoms; therefore, the frequency resolution can be varied. A wide variety of other wavelet-based harmonic dictionaries that involve various space-frequency partitions have been proposed in the recent years; examples are Gabor, curvelets and contourlets [40].

The atoms of a dictionary can also be created from the image itself. The first approach used in signal processing to create a dictionary from the image/signal is the Karhunen-Loeve Transform (KLT) [41]. An image consists of an array of pixels and there is some degree of correlation between the neighboring pixels of an image. KLT is the linear transformation which removes redundancies by decorrelating the image. KLT is generally computed from the covariance matrix of the image.

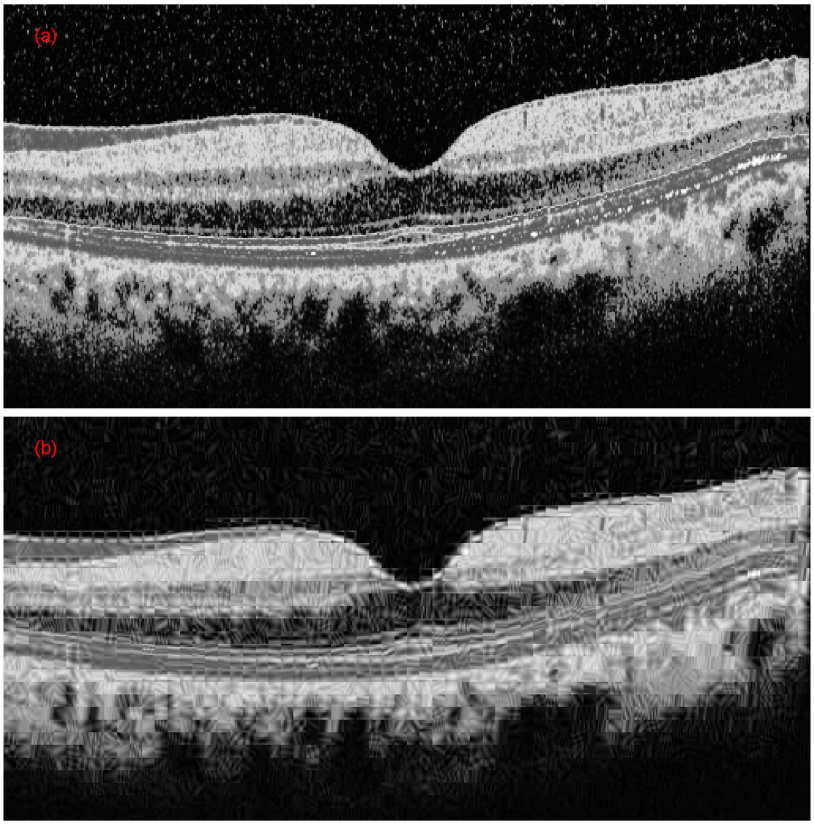


Figure 2.3: (a) Original Image (b) Image reconstructed using DCT dictionary. Blocking effect are seen in the reconstructed image

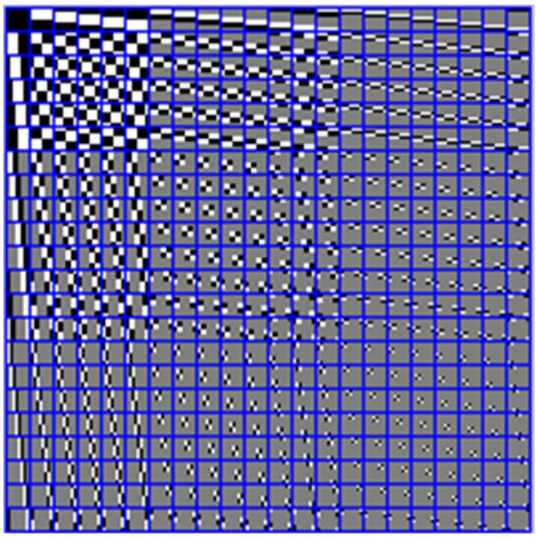


Figure 2.4: A Haar dictionary

Consider an n-dimensional signal $\{z_1, z_2, \cdots, z_N\}$. The KLT of any n-dimensional vector z is

$$y = V^T(z - \mu_z) \tag{2.8}$$

where $\mu_z = \frac{1}{N}\sum_{i=1}^{N} z_i$ is the mean vector of the given signal. The column vector of V is the eigenvector corresponding to the covarience matrix

$$C_v = \frac{1}{N}\Big(\sum_{i=1}^{N}[(z_i - \mu_z)(z_i - \mu_z)^T]\Big). \tag{2.9}$$

The basis functions of the KLT are the eigenvectors of the covariance matrix C_v [41]. Principal Component Analysis (PCA), which is closely related to KLT, is also widely used in dictionary design in sparse representation. The PCA uses an orthogonal transform matrix that converts a set of correlated variables into a new set of linearly uncorrelated variables called principal components. The transformation is designed in such a way that the variances of the principal components are arranged in decreasing order. Only the K most uncorrelated variables are kept; and therefore the dimensionality of the data is reduced. PCA decorrelates multivariate data, finds useful components and reduces dimensionality of the data. The PCA is obtained by using either the SVD of the Data Matrix, or eigenvalue decomposition of the covariance matrix [42].

Consider a signal $Z = [z_1, z_2, \cdots, z_M]$ in which $z_i \epsilon R^N$. The SVD of Z is: $Z = U\Sigma V^T$, where U is a $N \times N$ orthogonal matrix of eigenvectors of ZZ^T, V is a $M \times M$ orthogonal matrix of eigenvectors of Z^TZ and Σ is a rectangular diagonal matrix containing singular values $(\sigma_1, \sigma_2, \cdots, \sigma_n)$ in its diagonal. The covariance matrix created from the signal Z is

$$C_v = \frac{1}{N-1} Z^T Z \tag{2.10}$$

$$C_v = \frac{1}{N-1} V\Sigma U^T U\Sigma V^T$$

$$C_v = \frac{1}{N-1} V\Sigma^2 V^T.$$

This shows that the eigenvectors of C_v are the same as the right singular vectors of Z. Therefore the eigenvalues of C_v can be computed from the singular values of Z as

$$\lambda_j = \frac{\sigma_j^2}{N-1}. \tag{2.11}$$

This shows the relationship between the PCA and SVD. The principal components are columns of $U\Sigma$ matrix such that the signal Z is represented by the eigenvectors V as $Z = U\Sigma V^T$.

The atoms of a dictionary can also be trained directly from a training sample. This method is called dictionary learning [29]. The dictionary trained from a training sample is adapted to a given sample, and therefore it may better represent the signal. There are various ways of learning a dictionary. Here we describe the most famous one, called K-SVD algorithm. The algorithm is a direct extension of K-means clustering that is used to solve vector quantization (VQ) problem.

VQ is a lossy data compression method in which a good codebook is selected that provides the least mean square error (MSE) [29]. Consider a codebook matrix $C = [c_1, c_2, \cdots, c_k]$ each column represents the codeword (codevector). The signal is represented by the codeword such that $z_i = Cx_i$, where $x_i = e_j$ is a vector containing all zero entries except the one in the j^{th} position. The index j is selected according to the following criteria

$$\forall_{k \neq j} = \|z_i - Ce_j\|_2^2 \leq \|z_i - Ce_k\|_2^2. \tag{2.12}$$

This is sparse coding of z_i in which only one atom is used to represent the signal and the coefficient is forced to have unit value. The overall MSE is given by

$$E = \sum_{i=1}^{K} \|z_i - Cx_i\|_2^2 = \|Z - CX\|_F^2. \tag{2.13}$$

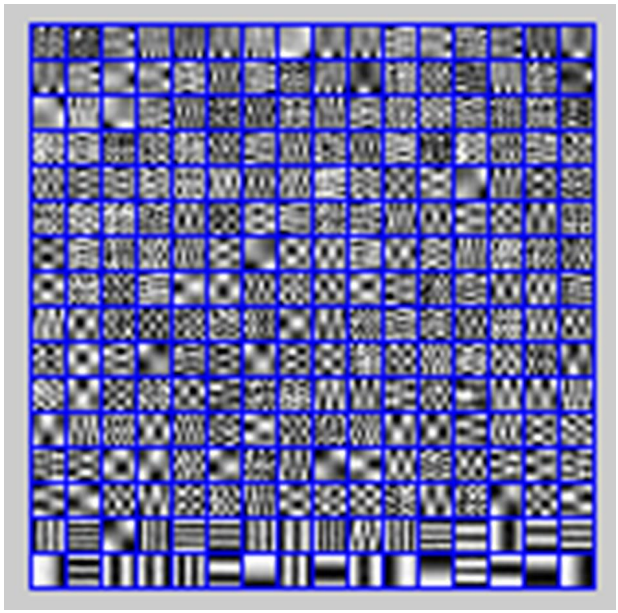


Figure 2.5: Adaptive dictionary trained from a fundus image using KSVD algorithm

The VQ training problem aims to train a codebook C that minimizes the error E, such that the coefficient vector contains only one non-zero entry and the value of the coefficients must be 1.

$$\begin{aligned} \min_{C,X} \quad & \|Z - CX\|_F^2 \\ \text{subject to} \quad & \forall_i, x_i = e_k \text{ for some k.} \end{aligned} \tag{2.14}$$

The above problem can be solved by using K-means algorithm. It is an iterative method that can be used to design an optimal codebook for VQ. The K-means algorithm comprises of two stages. The first stage is a sparse coding, in which a fixed codebook C is used to sparsely represent the signal while minimizing equation (2.14). The second stage updates the dictionary using fixed coefficient matrix X so as to minimize equation (2.14).

The K-SVD algorithm is considered as a generalization of the K-means algorithms in which instead of representing a signal with only one atom and fixing the coefficient to have the value of 1, the signal is represented by linear combinations of more than one atom and the coefficients can have any arbitrary values. For this case, minimization problem is

$$\begin{aligned} \min_{D,X} \quad & \|Z - DX\|_F^2 \\ \text{subject to} \quad & \forall_i, \|x_i\|_0 \leq L. \end{aligned} \tag{2.15}$$

Similar to K-means, this approach consists of two stages: sparse coding and dictionary update. In the sparse coding stage, a known dictionary D is used to determine the sparse vector X using any pursuit methods described in subsection (2.1.1). The dictionary update stage searches for a better dictionary. Since updating the overall dictionary D at the same time is impossible, the dictionary is updated one column at a time, keeping all the other columns fixed. For the column d_k of dictionary D, a new column $\tilde{d}_k$ and a new value for its coefficient that reduces the MSE in equation 2.15 are obtained. If we wish to update only one column d_k and its corresponding coefficient, the signal reconstruction error can be decomposed as follows

$$\|Z - DX\|_F^2 = \|Z - \sum_{j=1}^{K} d_j x_T^j\|_F^2 = \|\big(Z - \sum_{j \neq k} d_j x_T^j\big) - d_k x_T^k\|_F^2 = \|E_k - d_k x_T^k\|_F^2 \quad (2.16)$$

where E_k is the error associated to all the other atoms when d_k is removed. The process of updating only one column at a time can be achieved by using SVD; however, this process does not enforce the sparsity constraint. Therefore, the new vector x_T^k is most likely to be filled, which is against the sparsity constraint defined in equation 2.15. This problem can be tackled by using only a set of indices in X that uses the atom d_k that simply means indices of X where $x_T^k(i)$ is non-zero. Let w_k be such a set

$$w_k = \{i \mid 1 \leq i \leq K,\, x_T^k(i) \neq 0\}. \quad (2.17)$$

Now, construct a matrix $\Omega_k \epsilon R^{(N \times length(w_k))}$ with ones on the $(w_k(i), i)^{th}$ position and zeros elsewhere. The matrix transforms $x_R^k = x_T^k \Omega_k$ by keeping only non-zero coefficients, that converts the row vector x_T^k to the row vector x_R^k of length equal to $length \mid w_k \mid$. The same thing is applied to the input signal $Z_k^R = Z\Omega_k$ and error $E_k^R = E_k \Omega_k$, where Z_k^R represents the subset of input signal and E_k^R represents the portion of the error that are using the d_k atom. The minimization problem (equation 2.16) can be converted to give a solution vector $\tilde{x}_T^k$ that has the same support as the original x_T^k as follows

$$\|E_k \Omega_k - d_k x_T^k \Omega_k\|_F^2 = \|E_k^R - d_k x_R^k\|_F^2. \quad (2.18)$$

This can be easily solved using SVD method. SVD decomposes the error matrix $E_k^R = U \Delta V^T$. The first column of U is assigned as the new dictionary atom $\tilde{d}_k$ and its coefficient x_R^k is calculated by multiplying the first column of V by $\Delta(1,1)$. The same approach is followed to update all the columns of the dictionary and their corresponding coefficients in K-SVD dictionary learning approach. Figure 2.5 shows a trained dictionary from the fundus image.

2.2 Signal Recovery from Incomplete and Corrupted Measurements

In many signal processing applications, some entries of the data are missing or only a small set of data is accessible, or the data is corrupted by errors or noise [43]. The aim in such applications is to reconstruct the true set of data from an undersampled or corrupted data. There are many techniques to do this and they utilize some prior information of the data such as the sparsity, low-rankness and redundancy. This information is very important for efficient recovery of the data. The true signal can be recovered from the undersampled or corrupted signal by minimizing objective functions such as sparsity and matrix rank; however, these objective functions are in fact NP-hard. Recent methods utilize convex optimization that optimizes the convex relaxation of the original objective function under certain reasonable assumptions to give the optimal solution to the original problem [43]. In this section, two approaches are discussed that exploit convex optimization; one approach minimizes the sparsity of the signal and is called compressive sensing, and the other approach minimizes the rank of the matrix (rank minimization).

2.2.1 Compressive Sensing

A continuous signal of a finite bandwidth can be reconstructed accurately by a finite number of its discrete values. The number of discrete values is specified by the Nyquist-Shannon sampling theorem which states that a signal can be perfectly reconstructed when the sampling rate is set more than twice the highest frequency of the signal. Recently, many researchers have begun to investigate an alternative sampling paradigm called Compressive Sampling that goes against Nyquist-Shannon sampling in data acquisition. In this approach, under certain conditions a signal is uniquely reconstructed from samples far less than that required by Nyquist-Shannon theorem. The idea of reconstructing a signal from highly incomplete samples was proposed by Candes, Romberg, and Tao [44]; however, the name compressed sensing (compressive sampling) was coined by Donoho in 2006 [45]. The very first papers that introduced the concept of compressive sensing (CS) are [44, 45, 46].

Mathematically, consider a finite dimensional discrete time signal $z \epsilon R^N$. The signal can be sparsely represented in the transform domain. The coefficient vector has many zero coefficients and few non-zero coefficients if the signal is compressible. Commonly, a signal is said to be K-sparse in the transform domain if K of the N components of coefficient vector are non-zero. A signal is not sparse in the spatial domain; therefore, sampling less than twice of the Nyquist frequency yields aliasing. However, a signal is sparse in the

transform domain; therefore, K largest transform coefficients are sufficient to recover a signal without much perceptible loss. CS is a novel sampling paradigm that collects inner products between the signal and the basis vectors instead of directly sampling the intensities values from the scene. Let Φ be a collection of basis vectors and if we make $M << N$ inner products between the signal and basis vectors such as: $y_1 = \phi_1 z, y_2 = \phi_2 z, \cdots, y_M = \phi_M z$, then the CS data acquisition process is described by

$$y_{M\times 1} = \Phi_{M\times N} z_{N\times 1} \tag{2.19}$$

$$y_{M\times 1} = \Phi_{M\times N} \Psi_{N\times N} x_{N\times 1} \tag{2.20}$$

where $y_1, y_2, \cdots, y_M$ are the measured values. Equation 2.20 consists of two matrices: Φ for sampling the signal, and Ψ for transform coding. One of the crucial steps in CS is to carefully build these two matrices based on the specific properties of the signal. These two matrices should be highly incoherent. The coherence between any two matrices is within the range of $[1, \sqrt{N}]$. Intuitively, if the coherence between the two matrices is small, fewer samples are needed to reconstruct the original signal. In such case the signal can be estimated by only $M > K\log(N)$ linear measurements, which is far less than the actual bandwidth of the signal. For a K sparse signal if we make M linear measurements such that $M > K$, then the necessary and sufficient condition for signal recovery is the validity of the Restricted Isometry Property (RIP) for the measurement matrix Φ, i.e., the preservation of Euclidean distance. The RIP states that a matrix Φ has the RIP of order K if $\delta_k \epsilon (0, 1)$, such that

$$(1 - \delta_K) \leq \frac{\|\Phi z\|_2^2}{\|z\|_2^2} \leq (1 + \delta_K) \tag{2.21}$$

holds simultaneously for all sparse vectors z having no more than K non-zero entries [47]. For simplicity, by selecting entries of Φ from a random matrix or independent and identically distributed random variables from Gaussian or Bernoulli distributions, both the RIP and incoherence could be achieved with a high degree of probability, and the signal can be recovered from insufficient samples provided that the nonlinear signal recovery method is adopted [48]. For signal reconstruction the traditional least squares optimization is inadequate, so other types of signal reconstruction methods such as l_1-norm minimization or greedy algorithms must be invoked. For more information, we refer the readers to a recently published review paper by Thapa et al. [49].

2.2.2 Matrix Rank Minimization

Matrix rank minimization is the latest extension of CS. Matrix rank minimization is similar to the CS; however, instead of minimizing the l_0 norm of the sparse vector x of signal z (a

column of Z), it minimizes the l_0 norm of the singular value vector of Z. The low-rankness of a matrix requires that the number of columns or rows of the matrix to be much greater than the rank of the matrix. The rank of a matrix is equal to the number of its non-zero singular values [43] and can be calculated by taking the SVD of a matrix and arranging all the singular values into a vector. The matrix rank minimization method reconstructs the unknown matrix by utilizing its low-rankness. The aforementioned method has been widely used in the signal processing, such as, object recognition, collaborative filtering, matrix complication, etc.

Mathematically, consider a low-rank matrix $Z \epsilon R^{m \times n}; m \leq n$. Let $A : R^{m \times n} \rightarrow R^p$ be a linear transformation, and $b \epsilon R^p$. The rank minimization problem is defined as

$$\begin{aligned} \min_{Z} \quad & \text{rank}(Z) \\ \text{subject to} \quad & A(Z) = b. \end{aligned} \tag{2.22}$$

Here, the objective function is the rank of Z. This is an NP-hard problem so it is usually replaced by a computationally efficient method called nuclear norm minimization (NNM) $\|Z\|_* = \sum_{k=1}^{n} \sigma_k(Z)$, i.e., minimizing the sum of all singular values of Z

$$\begin{aligned} \min_{Z} \quad & \|Z\|_* \\ \text{subject to} \quad & A(Z) = b. \end{aligned} \tag{2.23}$$

The solution of equation (2.23) is exactly the same as the solution of equation (2.22) when it satisfied RIP condition for the matrix Z. For every integer r with $1 \leq r \leq m$, define isometry constant of A, δ_r to be the smallest number such that

$$(1 - \delta_r) \leq \frac{\|AZ\|_2^2}{\|Z\|_2^2} \leq (1 + \delta_r) \tag{2.24}$$

holds for all matrices Z of rank at most r [50].

There are various ways of solving rank minimization problem and utilizing SVD method is one of them. The low-rank approximation has several applications; however, in this thesis, we concentrate on its application for recovering true image from its noisy version. Consider a noisy matrix Y. The aim is to reconstruct a noise-free matrix Z from the noisy matrix Y by removing the noise. This is an inverse problem which seeks a true noise-free matrix from a noisy observation. There are several ways to solve such inverse problem. One way is to use NNM. That is, the noise-free matrix is obtained from the solution of the following energy function [51]

$$\hat{Z} = \min_{Z} \quad \|Y - Z\|_F^2 + \lambda \|Z\|_* \tag{2.25}$$

where $\|Z\|_* = \sum_i \mid \sigma_i(Z) \mid_1$ is the nuclear norm of the matrix Z; $\sigma_i(Z)$ is the i^{th} singular value of Z; $\|.\|_F$ is the Frobenius norm and λ is a positive constant. Cai et al. [52] proved that Eq. 2.25 can be solved using soft-thresholding operation on the singular values of the noisy matrix Y

$$\hat{Z} = US_\lambda(\Sigma)V^T \tag{2.26}$$

where U and V are orthogonal matrices and Σ is a diagonal matrix containing singular values $(\sigma_1, \sigma_2, \cdots, \sigma_n)$ in its diagonal. These matrices are obtained by taking the SVD of the noisy image Y (i.e., $Y = U\Sigma V^T$). The soft-thresholding function $S_\lambda(\Sigma)$ is performed on the singular values in such a way that for each diagonal element σ_i in Σ

$$S_\lambda(\sigma) = \max(\sigma_i - \lambda, 0). \tag{2.27}$$

The rationale for this approach is that the singular values are shrunk by soft-thresholding. The main downside with this approach is that, it shrinks all singular values by the same amount; therefore, we may lose important information from the matrix. In case of an image, it has been known that each singular value carries useful information about the image, but the larger ones carry more information than the smaller ones. Therefore, to preserve the important information, the larger singular values need to be shrunk less. Zhang et al. [53] introduced truncated NNM method that minimizes the smallest $n - r$ singular values, where n is the total number of singular values and r is the rank of the matrix Y. The practical challenge of this method is selecting the rank of the matrix. The rank of the matrix is very important since based on that the algorithm decides whether to regularize a specific singular value or not [51]. Recently, Gu et al. [51] proposed WNNM in which different weights are assigned to the singular values based on their magnitude. The weighted nuclear norm of the matrix Z is denoted by $\|Z\|_{w,*}$ and is defined as

$$\|Z\|_{w,*} = \sum_i \mid w_i\sigma_i(Z) \mid_1 \tag{2.28}$$

where $w_{i=1}^n \geq 0$ are the non-negative weights assigned to the singular values $\sigma_i(Z)$. Finally the noise-free image is reconstructed by utilizing these weighted singular values as: $\hat{Z} = UW_i(\Sigma)V^T$.

2.3 Image Quality Metrics

Many image quality metrics are used to test the performance of the image processing algorithms in this thesis. The most commonly used objective quality metric is the peak

SNR (PSNR). The PSNR is the peak signal-to-noise ratio between the predicted image and its original, in decibels. PSNR is calculated from the MSE, which is the average error between the predicted image and its original. Let $Z(i,j)$ be the original image and $\hat{Z}(i,j)$ be the predicted image. Then MSE and PSNR are defined as

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[Z(i,j) - \hat{Z}(i,j)\right]^2 \tag{2.29}$$

$$PSNR = 20\log_{10}\frac{L}{\sqrt{MSE}} \tag{2.30}$$

where L is the maximum signal value in the image. A higher value of PSNR indicates a better image quality. Structural similarity (SSIM) is another important object quality metric that computes the similarity between the predicted and original image. The SSIM accounts for the luminance, contrast, and structural changes between the two images. The SSIM index is defined as

$$SSIM = \frac{(2\mu_Z\mu_{\hat{Z}} + C_1)(2\sigma_{Z\hat{Z}} + C_2)}{(\mu_Z^2 + \mu_{\hat{Z}}^2 + C_1)(\sigma_Z^2 + \sigma_{\hat{Z}}^2 + C_2)} \tag{2.31}$$

where μ_Z and $\mu_{\hat{Z}}$ are the means and σ_Z^2 and $\sigma_{\hat{Z}}^2$ are the variances of the original and predicted images, respectively. $\sigma_{Z\hat{Z}}^2$ is the covariance of the predicted image and its original. $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$ are the two variables that stabilize the division when the denominator is weak; k_1 and k_2 are constants [54].

Cross-correlation (XCOR) is also calculated to estimate the quality of denoising. The XCOR is defined as [55]

$$XCOR = \frac{\sum_{i,j} Z(i,j)\,\hat{Z}(i,j)}{\sum_{i,j}[Z(i,j)]^2}. \tag{2.32}$$

XCOR measures the similarity between the two images. The XCOR value approaches to 1 when the predicted image is similar to its original.

Mean-to-standard-deviation ratio (MSR) and contrast-to-noise ratio (CNR) are the other two objective quality metrics used in image processing for examining the image quality. MSR and CNR are defined as follows

$$MSR = \frac{\bar{\mu_f}}{\sigma_f} \tag{2.33}$$

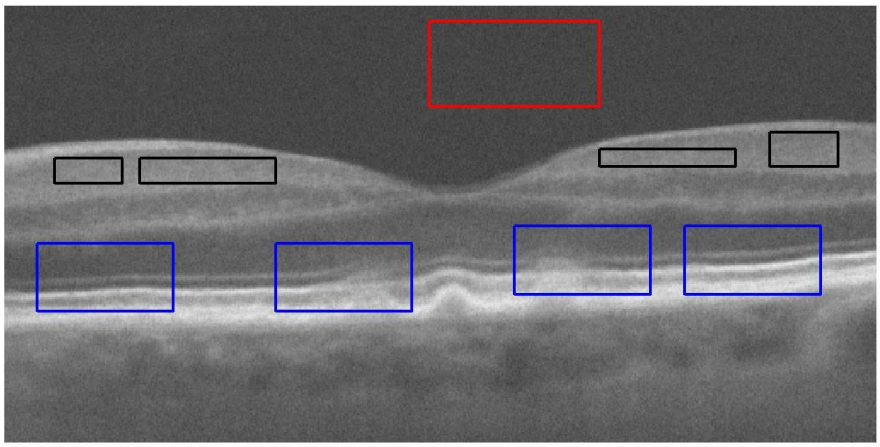


Figure 2.6: The image quality metrics are calculated from the foreground and background rectangular regions. MSR is calculated from homogeneous foreground regions (black colour rectangles) and CNR is calculated from both the homogeneous and inhomogeneous foreground regions (black and blue rectangles). The red rectangle represents the background region.

$$CNR = \frac{\mid \bar{\mu_f} - \bar{\mu_b} \mid}{\sqrt{0.5(\sigma_f^2 + \sigma_b^2)}} \tag{2.34}$$

where $\bar{\mu_f}$ and σ_f are the mean and standard deviation of the foreground region and $\bar{\mu_b}$ and σ_b are the mean and standard deviation of the background region [10]. The foreground and background regions of an image are shown in Figure 2.6. The red rectangle shows a background region and the black and blue rectangles show the foreground regions. The CNR measures the contrast between the foreground and background regions. MSR measures the smoothness of regions and is calculated from the regions that have homogeneous appearance.

Chapter 3

Retinal Image Reconstruction Using Nonlinear Dictionary

Chapter 2 highlighted a few examples of analytical-based and learning-based dictionaries. It has been found that the performance of analytical-based dictionaries highly depends on the nature of input signals. For example, DCT dictionary provides excellent results for smooth harmonic signals, but fails to provide similar results for transient signals. Similarly, wavelets, such as Haar and Gabor are well suited for transient signals; however, in comparison with DCT dictionary they do not perform well on harmonic signals. Nevertheless, the Gabor dictionary is unique in that it can be modified to work with both harmonic and non-harmonic signals by carefully adjusting the parameters of the function. However, selecting optimal parameters for generating appropriate Gabor atoms for an input signal is difficult [56]. On the other hand, learning-based dictionaries (adaptive dictionaries) perform well as they are created to best fit a given signal; however, the computational cost of training the dictionary is very high. Moreover, the dictionaries are adaptive so they do not work well with other signals of interest; therefore, a lengthy learning process is required. Adaptive dictionaries are adapted to a specific size based on the length of the signal and hence cannot be used for varying sizes of the signal. For example, if a dictionary is trained for a signal with 64 data points (i.e. 8×8 pixels patch of an image), it cannot accommodate for a signal with 49 data points (i.e. 7×7 pixels patch); indicating that the dictionary is not flexible. Another downside of adaptive dictionaries is that they require storing both the dictionary and sparse vectors. This consequently increases the storage requirement and hence decreases the compression efficiency. For online transformation both dictionary and sparse vectors need to be transferred which might be expensive. In contrast, the analytical-based dictionaries are better than adaptive dictionary in terms of

flexibility and storage requirement. Analytical-based dictionaries are simply mathematical functions and can be easily adjusted to accommodate signals of any size. However, when it comes to performance and compression ratio, the adaptive dictionaries are better than analytical-based dictionary. Many researchers have developed different types of dictionaries that result in better signal representation and research in sparsity-based signal processing continues. We have also created a hybrid dictionary that performs better than the known analytical-based dictionaries. In this chapter, we provide the performance of hybrid dictionary for retinal image reconstruction.

3.1 Nonlinear Atoms

3.1.1 Phase Added DCT Atoms

As discussed in chapter 2, DCT atoms are created from cosine function by changing the frequency; however, the phases are normally ignored. Therefore, a signal with a non-zero phase cannot be perfectly represented by the DCT atoms. In such a case, the signal reconstruction error is high. Marsousi [40] has shown that adding extra phase components to the DCT atoms improves the signal reconstruction. For each frequency in the horizontal and vertical direction, a set of evenly selected phases between 0 to 2π can be incorporated to the conventional DCT to create phase added DCT atoms. The phase can be chosen from the set $\phi\epsilon\{0, 2\pi/N_P, 4\pi/N_P, \cdots, 2(N_p-1)\pi/N_P\}$; where $N_p = 1, 2, 3, \cdots$ is the number of phase division between 0 to 2π. In 1-D, the phase added DCT atoms are generated using Eq. 3.1 where k represents the size of the discrete signal

$$d_k(t) = \cos\big[\frac{(2t+1)\pi k}{2N} + \phi\big], k = 0, 1, \cdots, N-1. \tag{3.1}$$

The aforementioned process increases the size of the dictionary. The phase added dictionary atoms created by adding phases $\phi\epsilon\{0, 2\pi/3, 4\pi/3\}$ are shown in Figure 3.1.

To demonstrate the performance of phase added dictionary, let us consider four different types of signals as shown in Table 3.1. The value of t was limited to -5 to 5 and the length of the signal was kept at 64. The OMP algorithm was used for signal coding. After coding, the signals were reconstructed using only 3 non-zero sparse coefficients. DCT and phase added DCT dictionaries were used for both sparse coding and signal reconstruction. The performances of dictionaries were compared by calculating the MSE and PSNR between the original and reconstructed signals. The results showed that MSEs were small when phase added DCT dictionary was used. Table 3.1 shows that the MSE decreases when phase

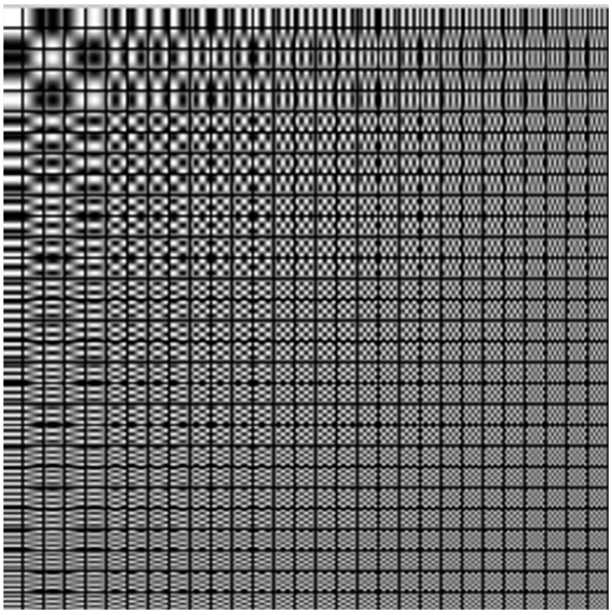


Figure 3.1: A phase added DCT dictionary.

Table 3.1: Performance of phase added atoms

	DCT		Phase added DCT	
Signals	**MSE**	**PSNR**	**MSE**	**PSNR**
$y_1 = 5\cos(2\pi 5t)$	0.0143	66.58	0.0133	66.90
$y_2 = 5\cos(2\pi 5t + \pi/2)$	5.2490	40.93	0.0129	67.04
$y_3 = 5\cos(2\pi 5t + \pi/3)$	2.4202	44.29	0.0123	67.25
$y_4 = 5\cos(2\pi 5t + 3\pi/2)$	2.5215	40.71	0.0129	67.04

added DCT atoms are used for reconstructing the signal with non-zero phases. Similarly, PSNRs were better when phase added dictionary was used for signal reconstruction.

3.1.2 Polynomial Atoms

The DCT atoms cannot precisely reconstruct a signal if it consists of non-harmonic components, such as polynomials. The accuracy of signal reconstruction can be increased by

Table 3.2: Performance of polynomial atoms

	DCT		DCT+Polynomials	
Signals	**MSE**	**PSNR**	**MSE**	**PSNR**
$y_5 = 5\cos(2\pi 5t)$	0.0954	58.33	0.0954	58.33
$y_6 = 5t^2 + 5\ \cos(2\pi 5t)$	47.69	31.35	0.2970	53.40
$y_7 = t^4 - 2t^2 + t$	2498.50	14.15	21.6205	34.78

adding a few polynomials atoms. The simplest polynomial atoms can be created using the following function $d_n(t) = t^n$; where t is linearly spaced points equal to the length of the vector and $n = 1, 2, \cdots, P$ order of the polynomial. The low order polynomial atoms characterize the non-harmonic smooth signals and higher order polynomial atoms characterize the transient signals; therefore, including these atoms in the DCT atoms facilitates reconstruction of both non-harmonic and transient signals. To illustrate this, consider signals y_5, y_6 and y_7 as shown in Table 3.2 .The signal y_6 consists of both harmonic (cosine) and non-harmonic (non-cosine) polynomial parts. Twenty polynomial atoms were created using the function $d_n(t) = t^n$ in which the values of t were taken from equally spaced values between -2 and 2. After coding, the signal was reconstructed with OMP method using only 3 sparse coefficients vector. The signal reconstruction results are shown in Table 3.2.

The MSE with the DCT atoms alone was 47.69 for signal y_6. With the addition of polynomial atoms, MSE reduced to 0.2970. However, for the harmonic signal y_5, the polynomials atoms did not show any effect. The MSE for signal y_7 is 2498.50 when DCT atoms were used, and was reduced to 21.62 when DCT plus polynomial atoms were used. Similarly, PSNRs increased significantly when polynomial atoms were added to DCT dictionary for reconstructing signals with non-harmonic components. Inspired by these results, we generated more structured polynomials atoms by translating and dilating the basic polynomial function of the form as shown in Eq. 3.2

$$d_n(t) = t^{n-1} + \sum_{i=0}^{n-1} a_i t^i. \tag{3.2}$$

Equation 3.2 consists of a set of polynomial constants. These constants can be generated in many ways. However, care should be taken to generate the coefficients since these coefficients are generated and stored prior to the construction of the polynomial atoms. The aforementioned increases the storage requirement and computational cost of the dictionary. For example, if we generate 64 coefficients using Eq. 3.2, we need to store at least 64

coefficients. The increased storage requirement reduces the ability to compress the signal. In this thesis, the coefficients were generated by fixed spacing scheme such that $a_i = a_0 + \Delta i$ where $i = 1, 2, \cdots, k$, a_0 is the starting coefficient value, and Δ is a fixed incremental step. The coefficients were circularly shifted to create the coefficients for another polynomial atom. In this method, only the starting and final coefficient values and the incremental step are stored. This will reduce the storage requirement to some extent and also allows generating atoms from various polynomial functions without describing each polynomial. The coefficients can be used in Eq. 3.2 to create a set of polynomial atoms. Besides coefficients, the polynomial atoms can easily cross the boundaries for larger values of t; therefore, the value of t should be limited to a certain range. In this thesis, t was set to equally spaced values between -5 and 5.

3.1.3 Exponential and Logarithmic Atoms

One of the common dictionaries which use exponential functions is the Gabor dictionary. Exponential functions have been used for many image processing tasks. The exponential atoms can also be created from polynomials by taking the exponential value of the polynomial function. In this thesis, the exponential atoms are created using Equation 3.3

$$d_n(t) = \exp(P_n(t)) \tag{3.3}$$

where $P_n(t)$ are the generic polynomial atoms. Similarly, we can create logarithmic atoms by taking the log of the polynomials functions as shown in Eq. 3.4

$$d_n(t) = \log(P_n(t)). \tag{3.4}$$

3.1.4 Shifted Boxcar Atoms

A boxcar function is a function that has a constant value over a small interval and zero elsewhere. It is a simple step function that resembles a boxcar, hence the name. The boxcar function can be defined as

$$d_n(t) = \begin{cases} c & a \leq t \leq b \\ 0 & \text{else.} \end{cases} \tag{3.5}$$

The boxcar function can be thought of as an extension of the delta function. The delta function is non-zero at a single point whereas the boxcar function is non-zero for a small

range. The boxcar function can be used to describe small and transient signals. We created boxcar atoms by performing N-circular shifts on the elementary boxcar function of constant width. By changing the width of the function, we produced an over-complete boxcar dictionary. The boxcar dictionary performance is similar to the Haar wavelet dictionary; therefore, it performs poorly for smooth signal and strong for transient signals. The performance of the DCT dictionary can be increased by adding boxcar functions so that the combined dictionary represents both the harmonic and transient signals. In this thesis, the boxcar atoms were created by keeping $c = 1$ and varying the width from 2 to 4 data points in the atoms.

3.1.5 Proposed Nonlinear Dictionary

We created a NL dictionary by adding phase shifted cosine functions and NL functions, such as polynomials, exponential and boxcar atoms to the DCT dictionary. The hybrid dictionary consists of a diverse set of atoms that are able to describe smooth harmonic, non-harmonic and transient signals. Table 3.3 shows a list of atoms that are used to create NL dictionary. The inclusion of the NL atoms will increase the size of the dictionary and therefore increases the sparse coding process as compared to the DCT dictionary. However, NL dictionary does not require time consuming dictionary learning algorithm; so it is much faster and less expensive than the adaptive dictionary. To reduce the sparse coding time, we can reduce the size of the NL dictionary by removing duplicate and useless atoms from the dictionary. An example of dictionary optimization has been proposed by Tharmalingam [56].

Once the atoms are arranged they are scanned for discontinuities; the dictionary columns are normalized. This new dictionary can be used for various sparsity-based image processing techniques. We can also add other functions to the NL dictionary; however, the addition of new atoms increases both the size and complexity of the dictionary.

The NL dictionary was compared with Haar wavelet and DCT dictionaries for the following three different types of signals y_8, y_9 and y_{10} as shown in Table 3.3. The signals consist of both harmonic and non-harmonic components. For all three dictionaries, OMP algorithm was used for sparse coding and the signals were reconstructed from only 3 sparse coefficients. The NL dictionary showed least MSEs for all 3 signals y_8, y_9 and y_{10} compared to DCT and Haar dictionaries.

The results for the signals y_8, y_9 and y_{10} are also shown in Figures 3.2, 3.3 and 3.4, respectively. The total number of non-zero coefficients (sparsity constant) was fixed to 3, 6, 9, 12, and 15 for 5 different cases. The signal was reconstructed from the sparse

Table 3.3: Different atoms of NL dictionary

Functions	**Tunable Parameters**	**No. of Atoms**
Phase added DCT: $D_k(t) = \cos\left[\frac{(2k+1)\pi t}{2N} + \phi\right]$	$\phi\epsilon\{0, 2\pi/N_p, \cdots, 2(N_p-1)\pi/N_p\}$ where $N_p = 5$	321
Simple polynomials: $P_k(t) = t^k$	$-2 < t < 2$	20
Structured polynomials: $P_k(t) = t^{k-1} + \sum_{i=0}^{k-1} a_i t^i$	$a_i = a_0 + \Delta i$	30
Exponential atoms: $E_k(t) = \exp(P_k(t))$	$0 < t < 1$	11
Boxcarl atoms: $B_k(t) = \begin{cases} 1 & a \le t \le b \\ 0 & \text{else} \end{cases}$	widths=$\{2:3\}$	130

Table 3.4: Performance of DCT, Haar and NL dictionaries for sparse signal reconstruction.

	DCT		Haar		NL	
Signals	**MSE**	**PSNR**	**MSE**	**PSNR**	**MSE**	**PSNR**
$y_8 = \cos(7t)$	0.0082	69.00	0.27	53.81	0.0011	77.73
$y_9 = 2t^2 + 5\cos(2\pi 5t)$	18.31	35.50	57.77	30.51	0.4618	51.48
$y_{10} = t^3 + 2t + 5\cos(2\pi 5t)$	45.09	31.58	126.77	27.10	4.911	41.22

coefficient vector. PSNRs were calculated to examine the performance of the dictionaries. The results showed that the NL dictionary had better performance for signal that consists of non-harmonic components.

3.2 Fundus and OCT Image Reconstruction using Nonlinear Dictionary

In this subsection, we describe the application of NL dictionary for retinal image reconstruction. Two fundus images (Figure 3.5) and two OCT images (Figure 3.6) were used

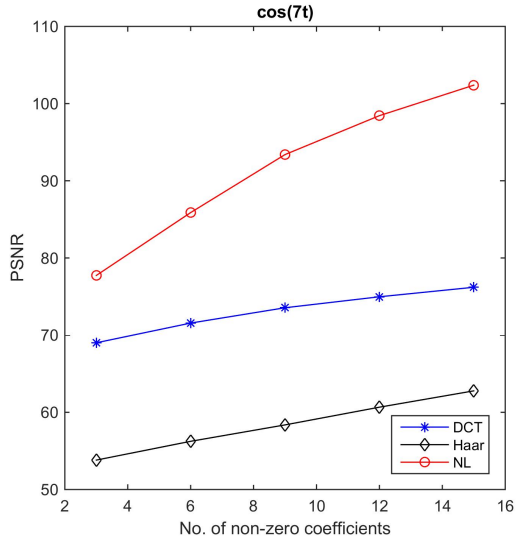


Figure 3.2: Performance comparison of DCT, Haar and NL dictionaries for $y = \cos(7t)$.

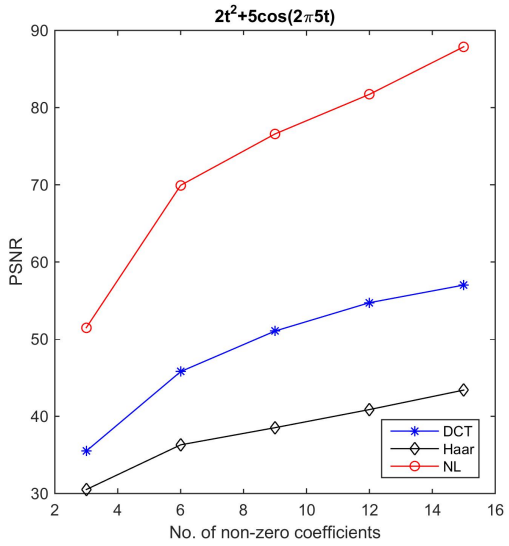


Figure 3.3: Performance comparison of DCT, Haar and NL dictionaries for $y = 2t^2 + 5\cos(2\pi 5t)$.

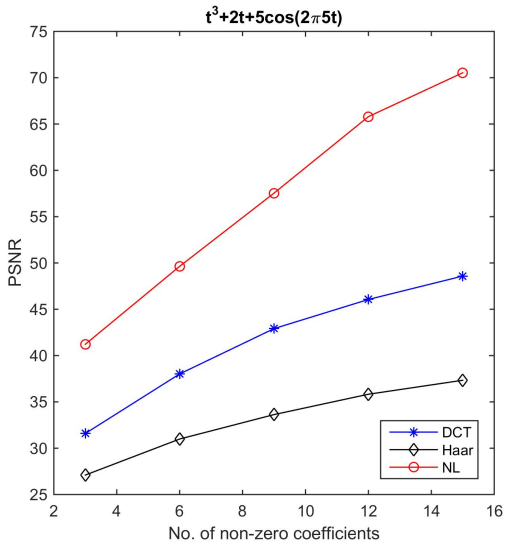


Figure 3.4: Performance comparison of DCT, Haar and NL dictionaries for $y = t^3 + 2t + 5\cos(2\pi 5t)$.

for the simulations. Using a fundus camera (Non-Mydriatic Auto Fundus Camera, Nidek AFC-230, Japan) fundus images were taken from the right eye of the author who has no ocular pathology. OCT images from two patients were provided by Sankara Nethralaya Eye Hospital Chennai, India.

3.2.1 Comparison Between Different Set of Nonlinear Atoms

To run the simulation, the images were broken down into 8×8 patches. Each patch was converted into one column vector. OMP algorithm was performed on each patch separately for sparse coding. The total number of non-zero coefficients (sparsity constant) was fixed to 3, 6, 9, 12, 15, 18, 21, 24, 27, and 30 for ten different cases. Finally, each patch was reconstructed from the sparse coefficient vector. The patches were reshaped to their square 8×8 shape and the image was reconstructed from the patches. To test the effect of different sets of atoms, the sparse coding was performed separately with (a) DCT atoms (b) phase added DCT atoms (c) phase added DCT atoms plus polynomials and exponential atoms, and (d) phase added DCT atoms plus polynomials, exponential and boxcar functions. PSNR and SSIM indices between the original and reconstructed images

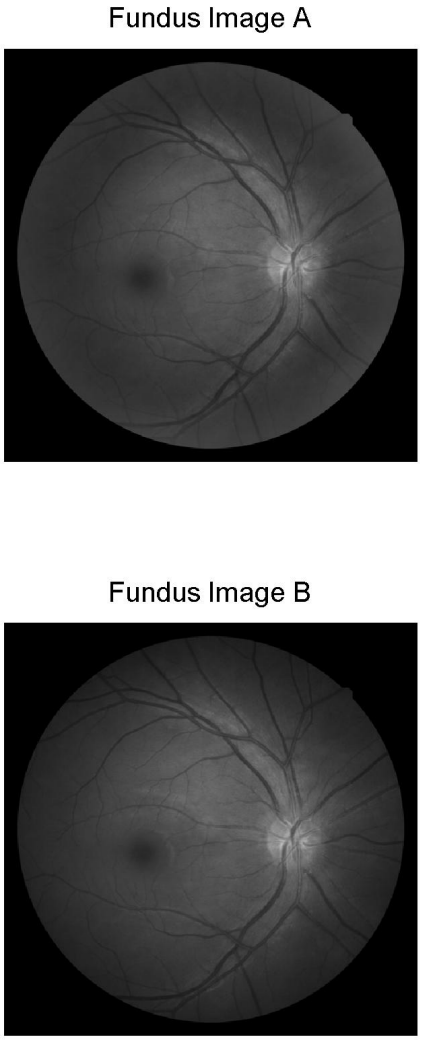
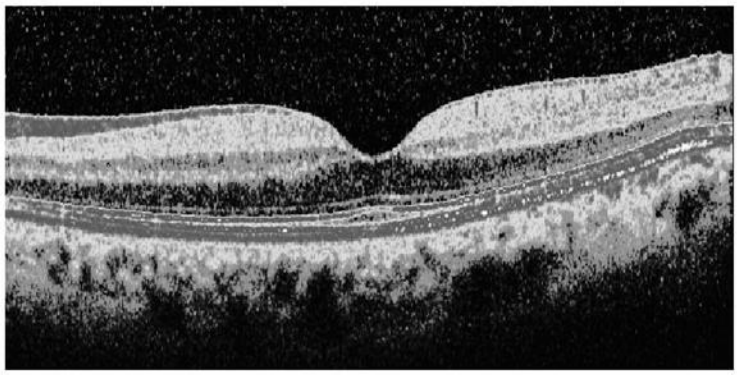


Figure 3.5: Test fundus images.

OCT Image A



OCT Image B

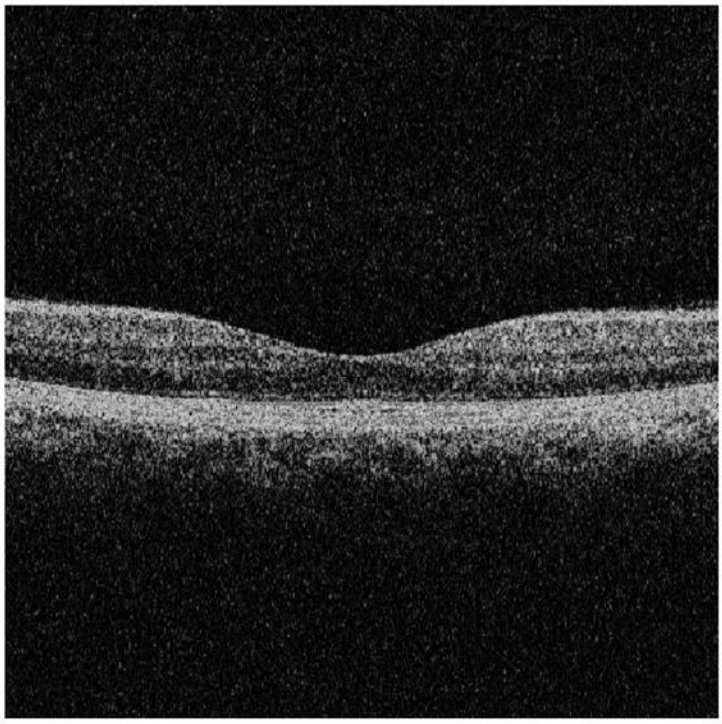


Figure 3.6: Test OCT images.

were calculated. MATLAB software (version R2013a) was used to code and implement the reconstruction process. A computer with the operating system 64 bit, Windows 7, Intel (R) Pentium (R) CPU G620T 2.2 GHz processor, and 4GB RAM was used. Figure 3.7 shows plots between PSNRs and number of non-zero coefficients. The results showed that PSNR improves with the addition of each of the aforementioned types of atoms. The improvement in PSNRs with the addition of polynomials, exponential and step functions was fairly noticeable for both fundus and OCT images. For example, there was at least 1.0 db improvement in the PSNR when fundus image A was reconstructed by 6 non-zero coefficients per patch (size of a patch 8×8 pixels) using a dictionary that was constructed by adding polynomial, exponential, and step functions to the DCT atoms. The improvement was more perceptible when the images were reconstructed with greater number of non-zero coefficients. For the same image, there was at least 5.0 db improvement in PSNR when the image was reconstructed by 30 non-zero coefficients; there was at least 3.5 db improvement in the PSNR when only phase, polynomial and exponential atoms were added to the DCT atoms. Figure 3.8 shows plots between SSIMs and the number of non-zero coefficients. Like PSNR, there was a clear perceptible improvement in the SSIM indices due to the addition of the polynomial, exponential and step functions for all fundus and OCT images. The dictionaries were the same size (64×512), therefore, the improvement is not due to the increased size of the dictionary.

3.2.2 Comparison with Other Known Dictionaries

The NL dictionary was compared with other dictionaries such as DCT, Haar and KSVD. The MATLAB source codes for KSVD dictionary learning were downloaded from Ron Rubinstein's webpage [57]. Similar to the previous section, the images were broken down into 8×8 patches and sparse coding was performed on each patch separately using OMP algorithm. The total number of non-zero coefficients (sparsity constant) was fixed to 3, 6, 9, 12, 15, 18, 21, 24, 27, and 30 for 10 different cases. Finally, the image was reconstructed from the sparse coefficients. The PSNR and SSIM indices calculated from the fundus and OCT images were plotted against the number of non-zero coefficients and are shown in Figures 3.9 - 3.12. The results showed that DCT, Haar, KSVD and NL dictionaries all improved PSNR when the number of non-zero coefficients per patch was increased.

The PSNRs and SSIMs obtained by NL dictionary were better than those obtained by DCT and Haar dictionaries; however, they were slightly less than those obtained from learned dictionary when fewer non-zero sparse coefficients were used. Nonetheless, the NL dictionary performed better than all three dictionaries when a large number of non-zero

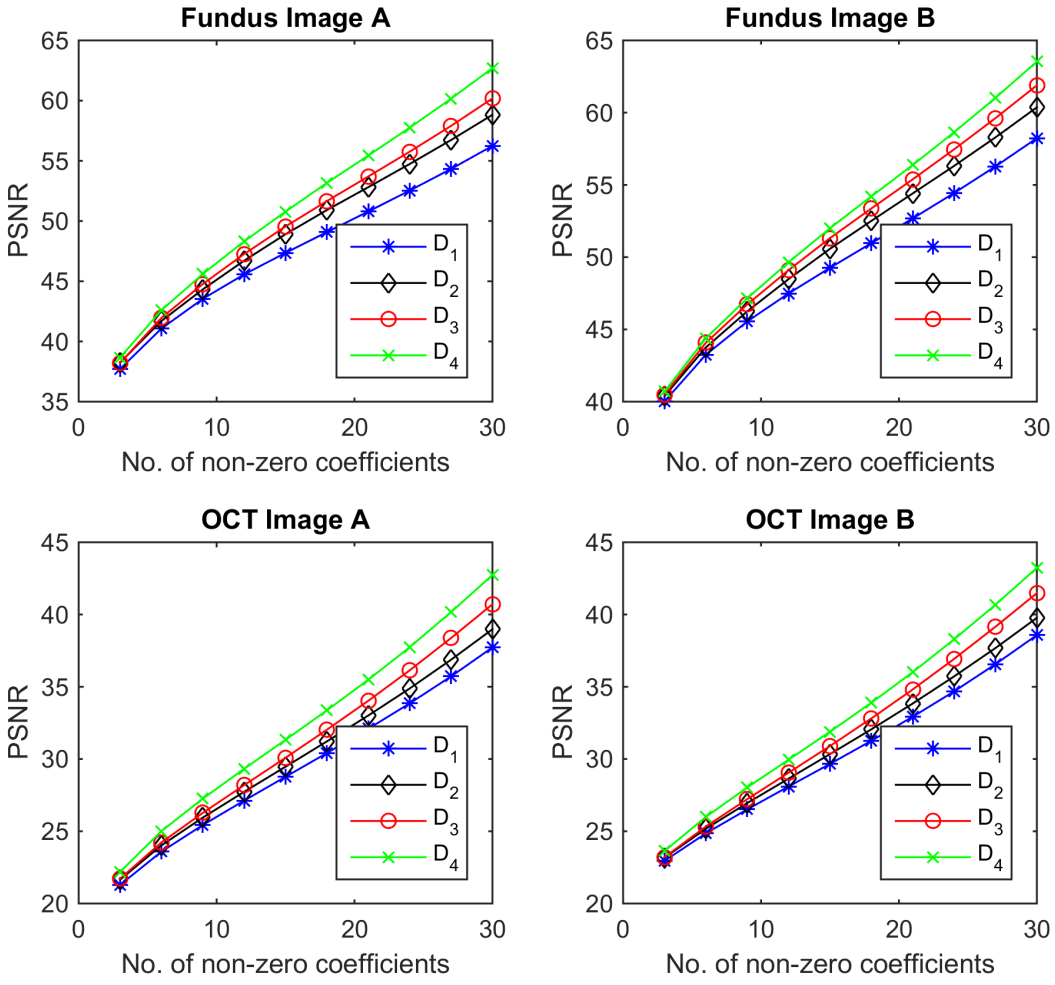


Figure 3.7: PSNRs obtained from each set of atoms of the NL dictionary. D_1=DCT; D_2= Phase added DCT; D_3=Phase added DCT plus polynomials and exponential functions; and D_4= Phase added DCT plus polynomials, exponential and step functions.

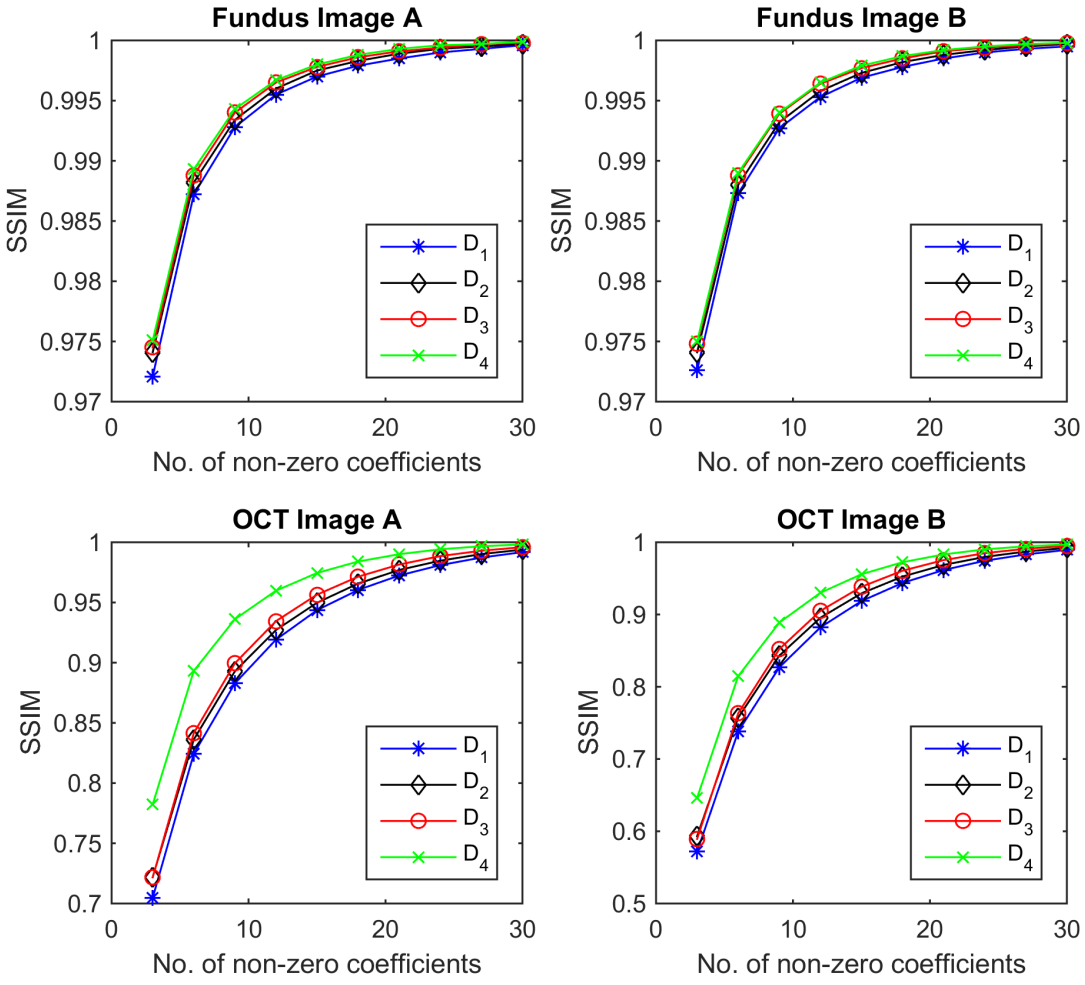


Figure 3.8: SSIM indices obtained from each set of atoms of the NL dictionary. D_1=DCT; D_2= Phase added DCT; D_3=Phase added DCT plus polynomials and exponential functions; and D_4= Phase added DCT plus polynomials, exponential and step functions.

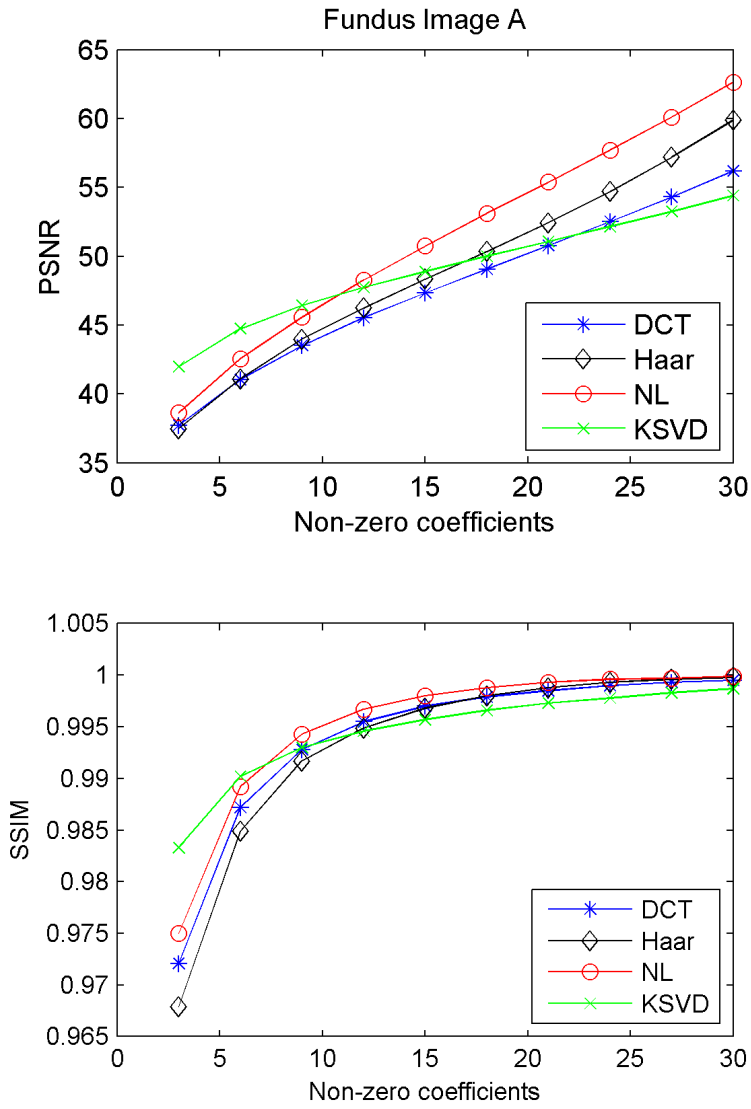


Figure 3.9: Performance of DCT, Haar, NL, and KSVD dictionaries for fundus image A reconstruction.

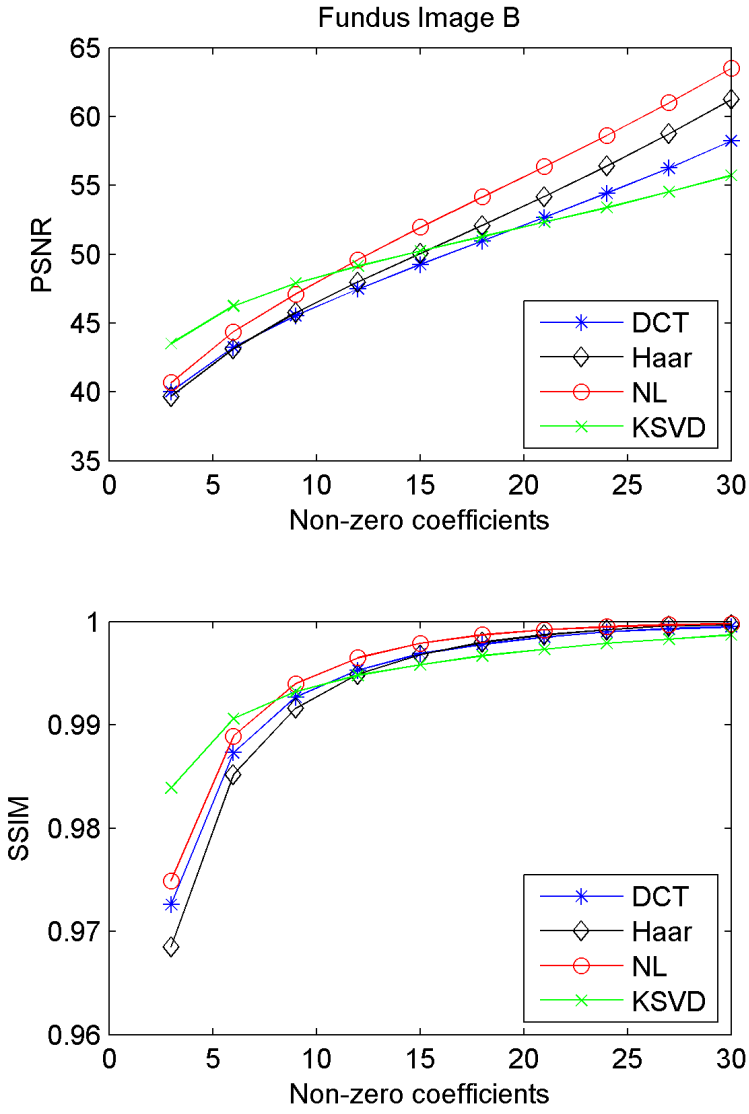


Figure 3.10: Performance of DCT, Haar, NL, and KSVD dictionaries for fundus image B reconstruction.

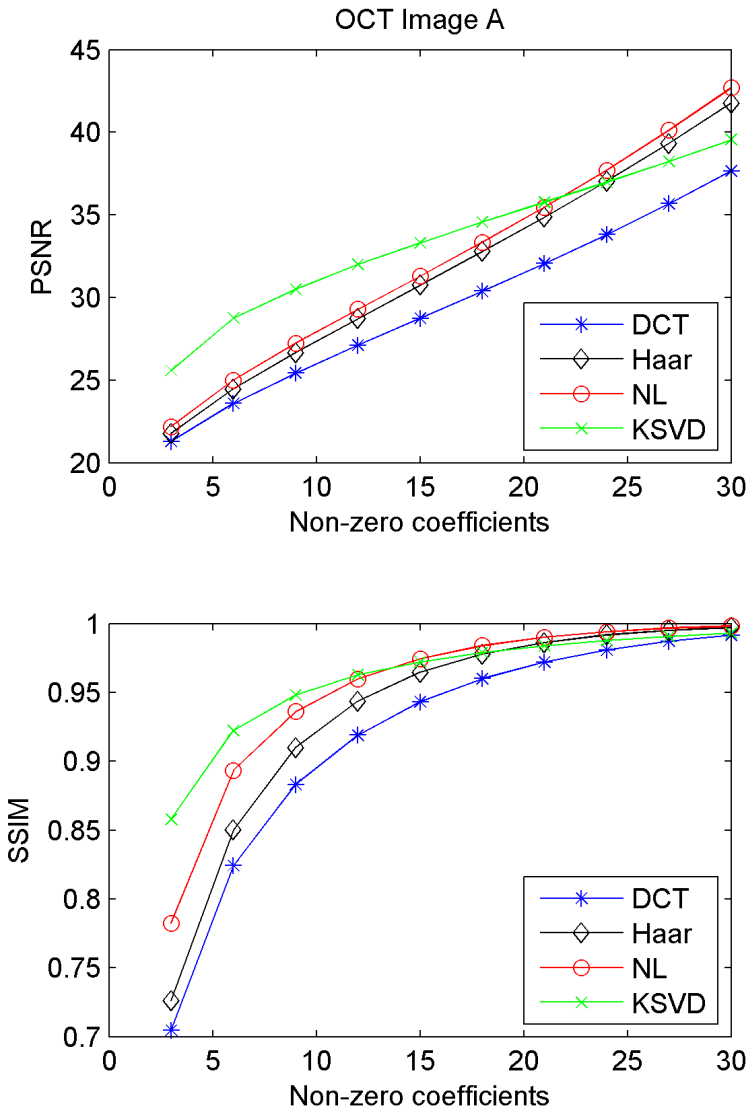


Figure 3.11: Performance of DCT, Haar, NL, and KSVD dictionaries for OCT image A reconstruction.

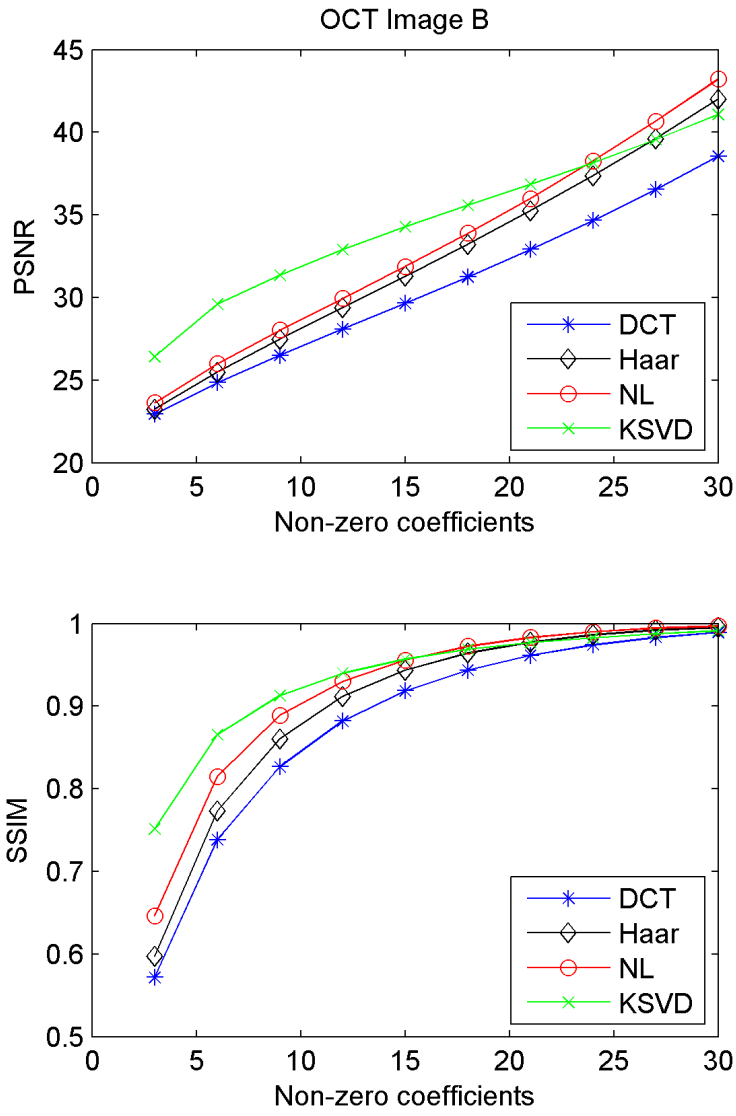


Figure 3.12: Performance of DCT, Haar, NL, and KSVD dictionaries for OCT image B reconstruction.

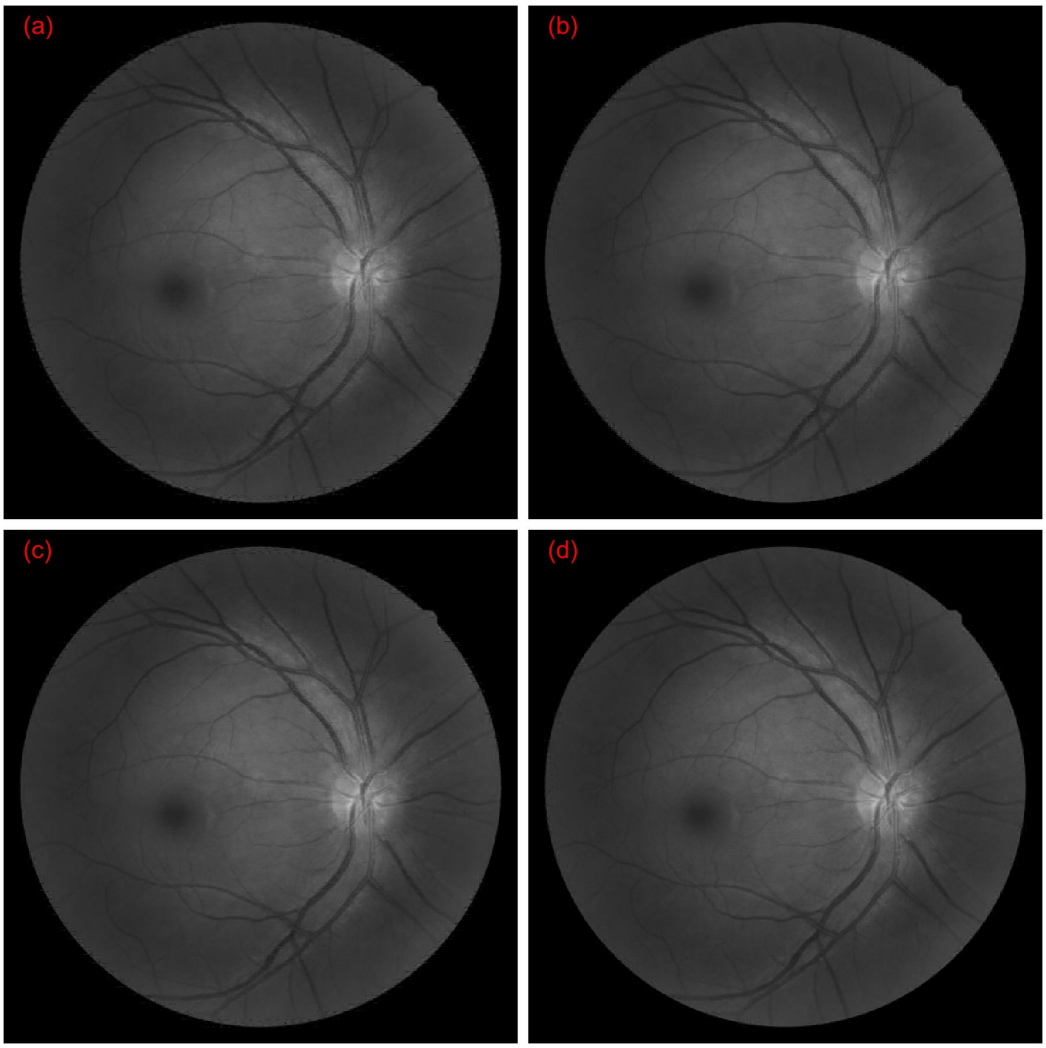


Figure 3.13: Fundus image A reconstructed from 6 non-zero coefficients per patch using (a) DCT (PSNR=41.05), (b) Haar (PSNR=41.09), (c) NL (PSNR= 42.58) and (d) KSVD (PSNR=44.68) dictionaries.

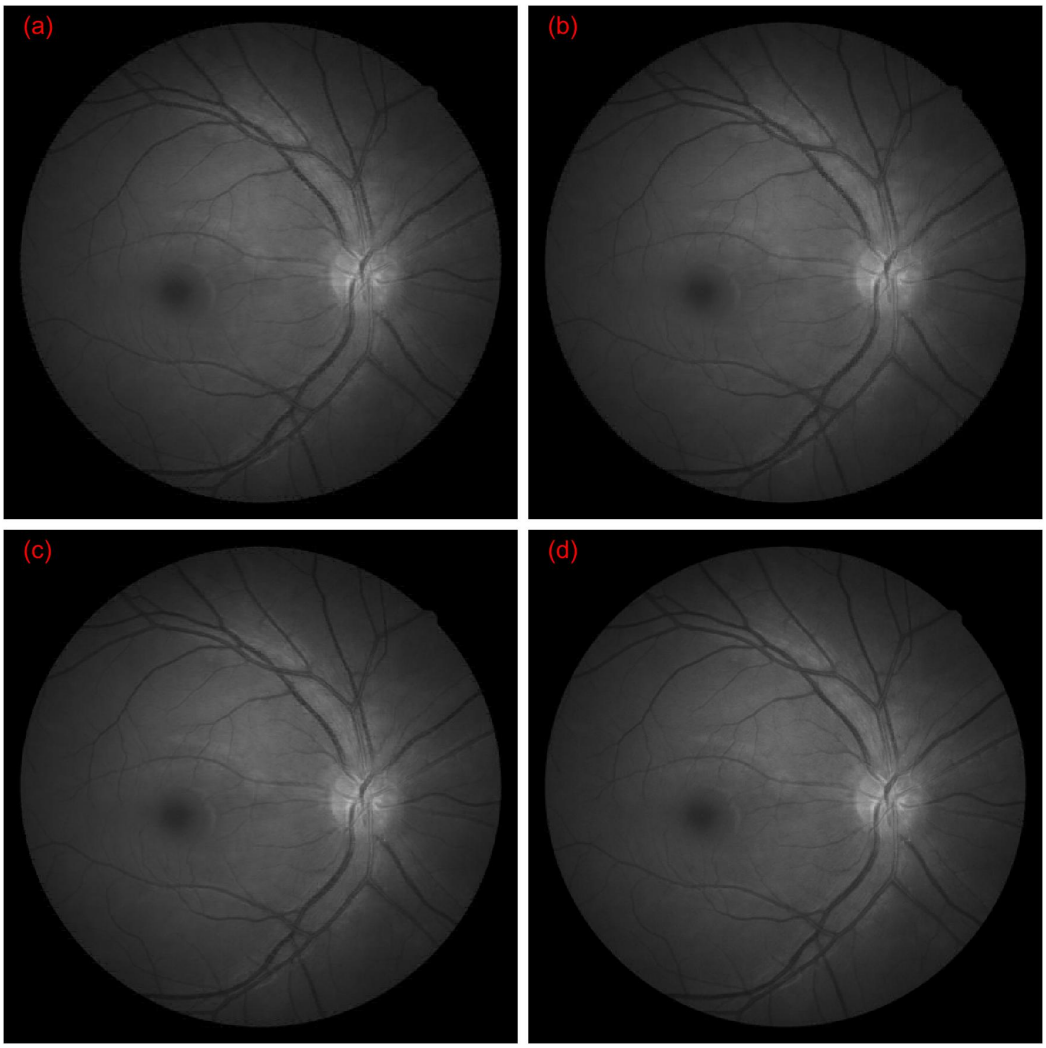


Figure 3.14: Fundus image B reconstructed from 6 non-zero coefficients per patch using (a) DCT (PSNR=43.24), (b) Haar (PSNR=43.12), (c) NL (PSNR= 44.37) and (d) KSVD (46.26) dictionaries.

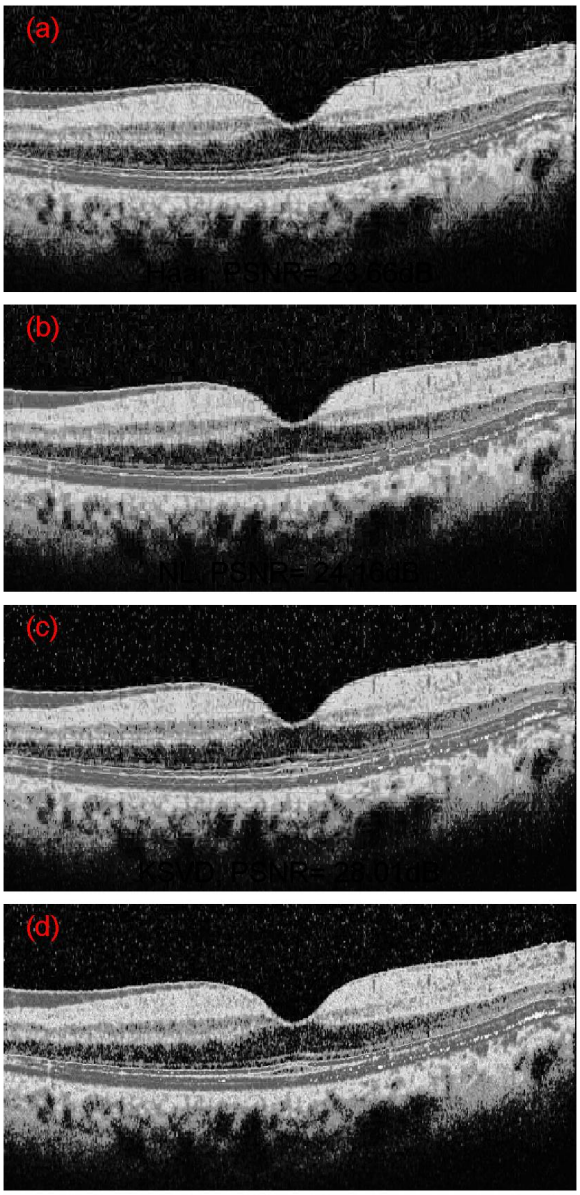


Figure 3.15: OCT image A reconstructed from 6 non-zero coefficients per patch using (a) DCT (PSNR= 23.59), (b) Haar (PSNR=24.47), (c) NL (PSNR=25.01) and (d) KSVD (PSNR=28.65) dictionaries.

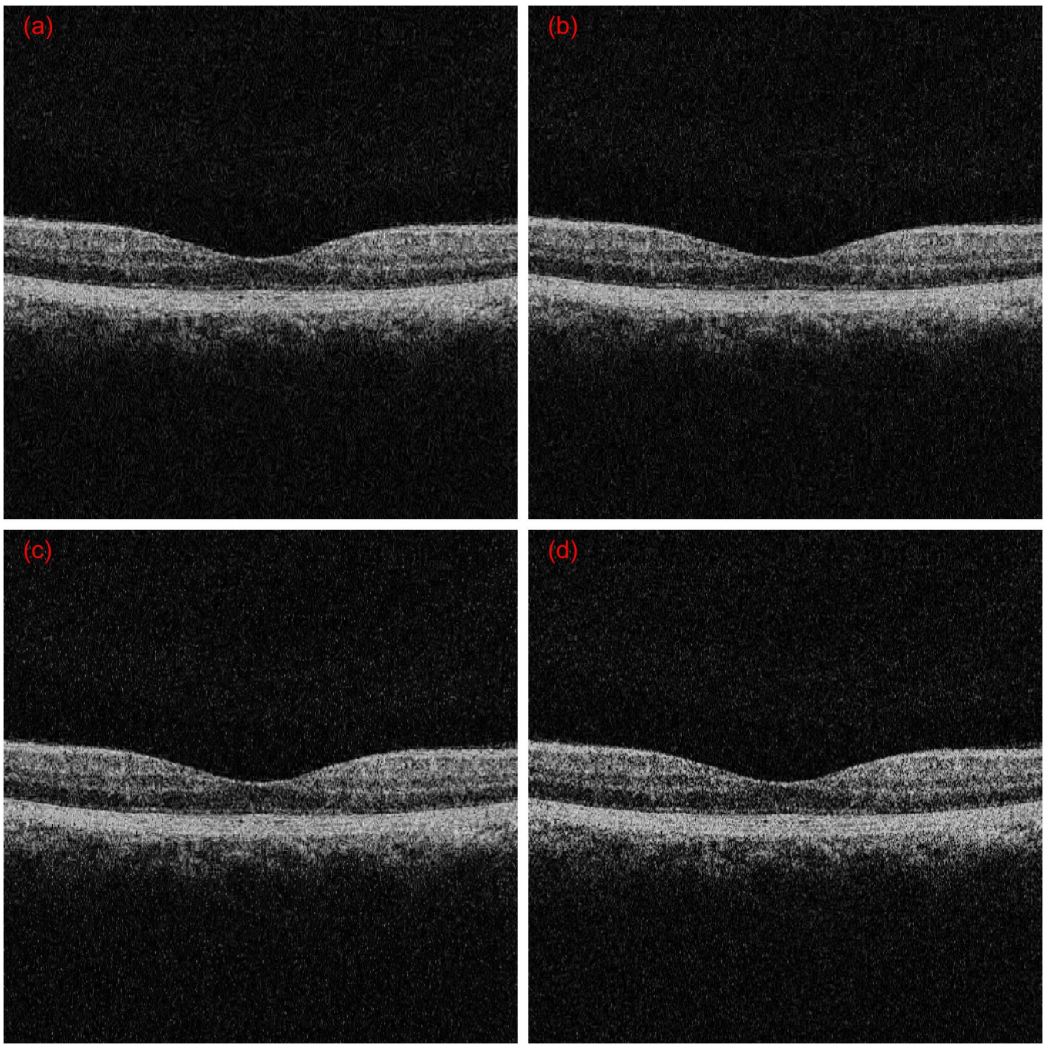


Figure 3.16: OCT image B reconstructed from 6 non-zero coefficients per patch using (a) DCT (PSNR=24.85), (b) Haar (PSNR=25.48), (c) NL (PSNR=26.00) and (d) KSVD (PSNR=29.63) dictionaries.

sparse coefficients were used. To get the same PSNR, the NL dictionary requires fewer non-zero coefficients than DCT and Haar dictionaries, which increases both the compression ratio and reconstruction time. There is a cross over between the NL and KSVD dictionary when there were about nine non-zero coefficients (see figures 3.9 to 3.12). Figures 3.9-3.12 show that in analytical based dictionaries such as NL, as the number of non-zero coefficients increased, the PSNR increased too; i.e., the plots showing changes in PSNR with changes in the number of non-zero coefficients are almost linear. However, the performance of the KSVD dictionary does not increase linearly as the number of non-zero coefficients increased. One reason for such finding is the effect of noise on KSVD dictionary. The KSVD dictionary is trained from the test images; therefore, some of its atoms represent the noise. The analytical dictionaries however, are created from a known function; therefore, their atoms are not influenced by noise.

Figures 3.13 -3.16 show fundus and OCT images reconstructed from 6 non-zero coefficients per patch using DCT, Haar, NL and KSVD dictionaries. The visual quality of the images reconstructed by NL dictionary was better than that of the images reconstructed by DCT and Haar dictionaries. However, the visual quality of the images reconstructed by NL dictionary was slightly poorer than that of the images reconstructed by KSVD dictionary.

3.3 Discussion and Conclusion

This chapter compared NL dictionary with various over-complete dictionaries in terms of image quality metrics, such as PSNR and SSIM indices. The experimental results showed that the performance gain is due to the addition of NL functions, such as polynomials, exponential, step functions and phase shifted cosine functions. The conventional DCT atoms provide excellent results for the smooth harmonic signals, but the retinal image may also contain some non-harmonic signals that cannot be perfectly represented by the DCT atoms. The NL dictionary includes a diverse set of atoms, which can reconstruct both the harmonic and non-harmonic signals; however, the inclusion of the atoms increases the computation cost. The computation time required to construct DCT dictionary was 0.207 seconds whereas the time required to create NL dictionary was 0.854 seconds which is four times greater than that of the DCT dictionary when compared in a computer with the operating system 64 bit, Windows 7, Intel (R) Pentium (R) CPU G620T 2.2 GHz processor, and 4GB RAM. This indicates that, NL dictionary is slightly more expensive than DCT when computing the same number of non-zero coefficients from sparse coding algorithm; however, the NL dictionary requires fewer number of non-zero coefficients to provide the same image quality (such as PSNR) so it is faster than DCT dictionary. The

NL dictionary is composed of diverse set of atoms; therefore, when constructed it requires more parameters compared to DCT dictionary, which increases the storage requirement. Nevertheless its storage requirement is far less than that of the adaptive dictionary. Since NL dictionary consists of many parameters, it is really difficult to implement it in hardware based applications such as cameras, micro-controllers and microprocessors. The suitable areas of application for the NL dictionary are those in which the compression ratio is more important than the compression time, such as medical image compression and storage, archives and CS. NL dictionary can also be used in telemedicine; the smaller data can be transmitted faster and the increased transmission capacity reduces the transmission cost.

Although NL dictionary provides slightly lower PSNR and SSIM indices than KSVD dictionary, it has three advantages over the adaptive dictionary. First, it does not require training set images therefor can be used in the real time situations. The second advantage involves data compression. If the data is compressed by learning-based dictionary, both the dictionary and sparse coefficients need to be stored, and therefore, it would take a large amount of computer space. For online transformation, both trained dictionary and sparse-coefficients need to be sent. However, if a signal is represented by an analytical-based dictionary, only sparse coefficients need to be sent or stored. This might have applications in telemedicine. Third, the construction of adaptive dictionary takes longer than that of the analytical dictionaries since it involves a learning algorithm that selects one dictionary atom at a time. The NL dictionary can replace adaptive dictionary if shorter execution time is required.

3.4 Summary

This chapter presented NL dictionary by adding various types of NL atoms such as phase added DCT, polynomials, exponential and steps functions to the conventional DCT atoms. The NL dictionary showed better performance when the signal consisted of both the harmonic and non-harmonic components. The NL dictionary was applied to reconstruct fundus and OCT images. The efficiency of image reconstruction increased with the addition of each sets of NL atoms. The NL dictionary performed better than DCT and Haar dictionaries for retinal image reconstruction; however, it showed slightly poorer performance than adaptive dictionary.

Chapter 4

Multi-frame Sparsity-based SD-OCT Image Denoising

Image denoising is a process of estimating a noise-free image from a noisy image or a series of images. Even though several methods have been developed to remove speckle noise from OCT images, image denoising still remains a challenging problem and an active field of research in image processing. The various denoising methods differ in terms of techniques, the noise model, and the amount of noise present in the image. OCT image denoising methods can be broadly divided into two groups: single-frame and multi-frame methods [58]. Single frame methods exploit either a standard filter or a sophisticated regularization method. Many single-frame methods have been developed for reducing the speckle noise; however, they either reduce the image resolution or increase the computational complexity, or both [59]. On the other hand, multi-frame methods take multiple images of the same or nearby locations and obtain a noise-suppressed image by averaging these images. In this chapter, we review the most commonly used speckle reduction methods and propose a new method based on the sparse representation algorithm.

4.1 Single-frame Denoising Methods

Filtering is the traditional approach for removing noise from an image. It can be done in spatial as well as transform domains. The most common spatial domain filtering methods include: mean filters, Wiener filters, median filters, Lee filters, Kuan filters, Geometric filters, Box filters, Gamma Maximum A Posteriori (MAP) filters, diffusion filters, non-local means, total variation (TV) regularization, and kernel regression. Transform domain

filtering methods are also broadly divided into the following two categories: data adaptive and non-data adaptive. If a transform matrix is constructed from the pre-specific set of functions, e.g. DCT and wavelet transform, then the filtering is called non-data adaptive. Wavelet domain methods are the most frequently used non-data adaptive transform domain methods for denoising images. This is because of their sparsity and multiresolution structures [58]. On the other hand, data adaptive filtering methods use a basis that is created from training set images, such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), KSVD and BM3D. We review both spatial domain filtering methods and transfrom domain methods in this subsection.

4.1.1 Spatial Domain Filtering

There are various types of spatial filters; the simplest of them is the Box filter [60]. The Box filter is a simple low pass filter that prevents the passage of high frequency noise components. However, it also smoothens the high frequency image components such as edges and fine details. Therefore, Box filter is less frequently used for denoising medical images. Another important filter used to remove speckle noise is median filter [60]. Median filter is created by constructing a window from an odd number of samples. The window selects a small section (equal to the size of the window) of the image and the middle pixel of this section is replaced by the median value of its neighbours. This process removes the noise by removing the abruptly changing pixels from the image. However, it provides wrong values at the edge of the image, blurring the edges and details. A more efficient filter for removing speckle noise is the Lee filter [61]. Lee filter, also called minimum mean square error filter, is based on the linear speckle noise model. It produces an output pixel from the following formula

$$\hat{I}(x,y) = \bar{I}(x,y) + W(x,y)[I(x,y) - \bar{I}(x,y)] \tag{4.1}$$

where $\bar{I}(x,y)$ is the mean intensity within the window and $W(x,y)$ is the weighting function. The weighting function is adaptive and is calculated from the following equation

$$W(x,y) = 1 - \frac{C_n^2}{C_I^2 + C_n^2} \tag{4.2}$$

where C_I is the coefficient of variance of the noised image and C_n is the coefficient of variance of the noise. The weighting function approaches zero in the uniform region; therefore, the output pixel is just the simple average of the pixels within the window. However, the weighting function approaches unity at edges and hence slightly modifies the

edges pixel values. Although the filter preserves the edges, its performance near edges is not optimum because it ignores some speckle noise. Another filter that is similar to Lee filter is the Kuan filter [62]. Both Lee and Kuan filter calculates the weighting functions based on the equivalent number of looks (ENL) from the image. The ENL estimates the noise variance and based on that smoothing will be performed. A smaller ENL leads to higher smoothing applied to the image. Kuan filter is different from Lee filter only in terms of its weighting function. The weighting function is calculated by

$$W(x,y) = \frac{1 - \frac{C_n^2}{C_I^2}}{1 + C_n^2}. \tag{4.3}$$

Wiener filter is another commonly used adaptive filter for removing noise and blur from an image [63]. It minimizes the overall MSE. The Wiener filter in the Fourier domain is given by

$$W(u,v) = \frac{H^*(u,v)\, S_I(u,v)}{\mid H(u,v) \mid^2 \; S_I(u,v) + S_n(u,v)} \tag{4.4}$$

where $H(u,v)$ is the blurring filter and $S_I(u,v)$ and $S_n(u,v)$ are the power spectra of the original image and noise, respectively. The noise and power spectra can be estimated in various ways. The filter removes the noise based on the calculation of local variance from the image. If the local variance is small, the image is smoothed significantly; however, if the local variance is large, the image is smoothed slightly to preserve the edges and fine details.

The Frost filter is another adaptive filter that estimates the new pixel from a weighted sum of the intensity values within the moving kernel [64]. The kernel moves across the image and the weights are calculated based on the local statistics in the moving window. The weight decreases with distance from the filter centre and increases if the variance of the kernel increases

$$W = \exp[-DC_I(s_0) \mid s \mid] \tag{4.5}$$

where D is the exponential damping factor, $C_I = \sigma_I / \bar{I}$; $\bar{I}$ and σ_I are the mean and standard deviation, s_0 is the location of the processed pixel and $\mid s \mid$ is the distance from the pixel s_0. The parameters are measured within a fixed size window. The output pixel is calculated from the following formula

$$\hat{I}_i = \frac{\sum I_i W_i}{\sum W_i}. \tag{4.6}$$

The improved Lee filter, improved Kuan Filter and improved Frost filter are the modified versions of Lee, Kaun and Frost filters, respectively [60]. These filters classify the image

into three different regions: homogeneous, heterogeneous and isolated, and exploit different filters and weights for each region to enhance denoising.

Another standard filter for removing speckle noise is Gamma MAP filter [60]. The Gamma MAP filter assumes that the values of the filtered pixel are gamma distributed. Another method uses partial differential equations for removing speckle noise from the images. An anisotropic diffusion filter denoises an image using the diffusion equation, a partial differential equation [65, 66]. The anisotropic diffusion scheme is given by

$$\frac{\partial I(i,j,t)}{\partial t} = \mathrm{div}[g(\|\nabla I\|)\nabla I] \tag{4.7}$$

where $\|\nabla I\|$ is the gradient magnitude, and $g\|\nabla I\|$ is an edge stopping function. Perona and Malik [65] proposed two different edge stopping functions

$$g(\nabla I) = \frac{1}{1 + \left(\frac{\|\nabla I\|}{K}\right)^2} \tag{4.8}$$

$$g(\nabla I) = \exp\Big[- \Big(\frac{\|\nabla I\|}{K}\Big)^2\Big] \tag{4.9}$$

where K is an edge magnitude parameter. These functions remove the noise from the uniform area; however, they cannot preserve the edge information. They were modified later by Black et al. [66] for a better edge-stopping function

$$d_n(t) = \begin{cases} \Big[1 - \frac{(\|\nabla I\|)^2}{K^2}\Big]^2 & , \text{if } (\|\nabla I\|)^2 \leq K^2 \\ 0 & , \quad \text{else.} \end{cases} \tag{4.10}$$

In anisotropic diffusion filtering, if the gradient magnitude is greater than edge magnitude (i.e. $\mid \nabla I \mid > K$), then $g(\nabla I) \to 0$, and if $\mid \nabla I \mid < K$, then $g(\nabla I) \to 1$. Diffusion is carried out if $(\|\nabla I\|)^2 \leq K^2$.

The previous methods replace the intensity of a pixel with the average intensity of the nearby pixels. However, the performance of a denoising method can be further enhanced by averaging the remote pixels that have similar intensities. Non-local means is a method that averages similar non-local pixels to reduce the image noise [67]. In this method pixels with similar neighbours will have larger weight than those with different neighbors. In practice, the search for similar pixels is carried out in a large neighborhood. The whole image is divided into small patches and the resemblance among the patches is evaluated to calculate the weight. The non-local means filter is defined as

$$NLI(p) = \frac{1}{C(p)} \int f(d(B(p), B(q))I(q)dq \tag{4.11}$$

where $f(d(B(p), B(q))$ is the weight, $d(B(p), B(q))$ is Euclidean distance between the two patches $B(p)$ and $B(q)$ centre at pixels p and q, respectively. The intensity of the pixel p is calculated from the following function

$$\hat{I}_i(p) = \frac{1}{C(p)} \sum_{q \epsilon B(p,s)} I_i(q)\, w(p,q) \tag{4.12}$$

where,

$$C(p) = \sum_{q \epsilon B(p,s)} w(p,q).$$

$B(p,s)$ is a neighbourhood with $(2s+1) \times (2s+1)$ pixels centered at p. We can choose the size of the neighbourhood based on the amount of noise in the image. A larger neighborhood provides better denoising results by averaging many pixels with higher computational complexity. The Euclidean distance $d = d(B(p,r), B(q,r))$ between the patches $B(p,r)$ and $B(q,r)$ of size $(2r+1) \times (2r+1)$ within the neighbourhood is calculated as

$$d^2(B(p,r), B(q,r)) = \frac{1}{(2r+1)^2} \sum_{q \epsilon B(0,r)} (I(p+i) - I(q+i))^2. \tag{4.13}$$

The weight $w(p,q)$ is calculated by

$$w(p,q) = \exp\big[- \frac{\max(d^2 - 2\sigma^2, 0)}{h^2} \big] \tag{4.14}$$

where σ is the standard deviation of the noise and h is the filtering parameter depends on σ. In this approach each pixel value is estimated from the weighted average of the similar pixels.

4.1.2 Transform Domain Approaches

The transform domain methods are further divided to non-data adaptive and data adaptive methods. The non-data adaptive approaches use a set of pre-defined basis functions; such as Fourier transform, DCT, and wavelet transform. The Wavelet transform transforms a noisy image into a wavelet domain, and shrinks the wavelet coefficients. Before shrinking the coefficients, multiplicative speckle noises are converted to additive noise by taking the lagorithmic transformation. Since speckles are high-frequency components and have non-zero wavelet coefficients when the image is transformed in the wavelet domain, they can be suppressed by shrinking the wavelet coefficients. Thresholding or statistical modelling

is used to remove the noise and/or the relatively less important wavelet coefficients. After noise suppression, an inverse wavelet transform is performed to reconstruct the image into the spatial domain. Following these steps, several wavelet domain based image denoising methods have been proposed in the past few decades [68].

An image can be effectively denoised using linear filters such as Weiner filters, or non-linear thresholding methods such as Hard Thresholding and/or Soft Thresholding of the wavelet coefficients [69, 70]. Thresholding methods convert wavelet coefficients to zero if their magnitudes are less than the threshold and maintain (Hard Thresholding) or modify (Soft Thresholding) them if their magnitudes are greater than the threshold. Thresholding methods are further divided into non-adaptive and adaptive thresholding. The best known non-adaptive thresholding method is VISUShrink [71], where a universal threshold is used to reduce the noise. On the other hand, SUREShrink is the most common adaptive thresholding method and performs better than VISUShrink where a threshold is derived by minimizing Steins unbiased risk estimator [72]. Furthermore, a new SURE-based orthonormal wavelet image denoising approach that does not need any prior statistical model of the wavelet coefficients is presented in [62]. The downsides of thresholding method are that first of all it is difficult to find the appropriate threshold, and second the same threshold cannot be used for different resolutions.

Recently, several researchers have developed models of wavelet coefficients and applied them to denoised images corrupted by additive white Gaussian noise. The models are either deterministic or statistic. One of the most successful statistical models that have performed well for modeling the wavelet coefficients is Gaussian State Mixture (GSM) [73]. GSM models the neighborhoods of coefficients as a product of Gaussian Random Vector and an Independent Hidden Random Scalar Multiplier. The other important wavelet-based denoising method uses the Bayesian least squares estimate of the noise-free data using a conditional posterior sampling approach [74]. Recently, a probabilistic model called Product of Edgeperts (PoEdges) [75] which models the statistical dependencies between coefficients in a wavelet decomposed image, has been introduced. This model shows impressive results in the field of image denoising. Although wavelets show promising results for features changing in horizontal and vertical directions, they have limitations when it comes to features in other orientations and hence do not represent the features changing along the edges of the image effectively. Other wavelet based speckle denoising methods have been proposed in the literature such as curvelet and contourlet [76, 77]. Curvelet transform is an extension of the wavelet transform that provides better performance along curves. Curvelet transform performs better than other methods of speckle reduction.

The data adaptive approaches train a basis (dictionary) from the training set images that best matches a given signal/image. This subsection describes the most important

data-adaptive transform domain denoising approaches. One such approach utilizes thresholding of singular values using SVD [78]. In this method, first the test image is divided into rectangular overlapping patches. Then similar patches are grouped together using clustering methods such as K-means clustering. A cluster matrix is extremely redundant; the unknown noise-free cluster matrix can be estimated utilizing its low-rankness. This can be done by utilizing NNM which is performed by thresholding singular values. The signal variance is used to set an initial threshold and the threshold is updated iteratively. The noise is reduced by soft-thresholding of singular values, and the image is reconstructed by averaging the overlapping section of the image. A recently proposed method uses WNNM, in which smaller singular values are shrunken more and the larger singular values are shrunken less [51].

Another approach uses PCA to reduce image noise. PCA algorithms transform the image into PCA domain and then eliminate the noise and minor principal components and reconstruct the image from a few most significant principal components. Applying PCA directly to noisy images without data selection has resulted in noise residual and many visual artifacts in the processed image. A new method called Local Pixel Grouping-Principle Component Analysis (LPG-PCA) uses a moving window to group similar pixels for training a local PCA transformation matrix [79]. Similarly, other recent methods divide an image into rectangular patches and group the patches into many structural clusters and train a local PCA transformation matrix from each cluster. Each data cluster is denoised by the corresponding local PCA transformation matrix [80].

Another approach frequently used for removing noise is utilizing sparse representation technique. A very popular technique that utilizes sparse representation of a signal is the denoising technique proposed by Elad and Aharon [81]. In this technique, a highly overcomplete dictionary is trained from the noisy image patches using KSVD algorithm and is used to sparsely represent the image patches. Each patch of the image can be approximated as a linear combination of few dictionary atoms; therefore, the representation vector is sparse. Only the most significant sparse coefficients are used to reconstruct the image patch, hence the noise is suppressed from patch. A clean image is reconstructed by averaging the overlapping portion of the denoised patches. The KSVD approach uses the sparsity of the image; however, it does not use the self-similarity of the image. Dabov et al. [82] proposed a denoising method called BM3D by utilizing the self-similarity of the image. The method was considered a cutting edge approach in the past as it proved to be an extremely effective method in natural image denoising. The method consists of three steps. First, it searches for similar 2D blocks in the image using a block matching algorithm and stacks them together to create a 3D cube. Second, a 3D transform of the cube is taken and the noise is removed by shrinking the transform coefficients through Wiener filtering. Third, an inverse

3D transform is performed to reconstruct a noise suppressed image in the spatial domain. Local methods such as KSVD dictionary learning method train a dictionary of a basis function that sparsely represents an image. On the other hand, the nonlocal methods such as BM3D achieve the sparsity by clustering the similar blocks and shrinking the coefficients. Dong et al. [83, 84] incorporated these two methods into a unified variational framework and developed a denoising method. In this method, an image is divided into rectangular patches. These patches are divided into many structural clusters using block matching algorithm. A compact PCA dictionary is computed from each structural cluster and is used for sparse representation of the patches within the cluster. Finally, the image is reconstructed by averaging the overlapping portions of sparse reconstructed patches. Fang et al. [10] used a similar approach in which the PCA dictionary was replaced with KSVD based learned dictionary.

4.2 Multi-frame Denoising Methods

Multi-frame denoising methods average multiple images of the same or nearby locations to obtain a noise-suppressed image. Multi-frame averaging is done is two ways. In one method, the optical setup or scanning protocols of the OCT is modified to acquire multiple images in a way that the speckle pattern differs but the structure of the images remains the same [59]. These images are then averaged to acquire a noise-suppressed image. Various spatial compounding and frequency compounding methods are reported under this category [59]. While this method works well, it may increase the complexity of the data acquisition process. The second method of multi-frame averaging is by modifying the software. In this method various post processing algorithms are used to suppress the speckle noise.

In the past decade, several digital signal processing-based multi-frame image denoising algorithms, from simple averaging to complex registration and optimization-based averaging algorithms, have been developed to reduce speckle noise in OCT images. These algorithms either denoise a single frame at a time and then average multiple denoised frames, or average multiple frames and denoise the averaged frame. Ozcan et al. [85] did not find any significant difference between the two approaches quantitatively; however, the computation time was much shorter when only the averaged frame was processed. Sunder et al. [86] proposed a robust correlation algorithm to line up A-scans of a specific retinal location from 5 to 15 consecutively recorded B-scans using TD-OCT. A weighted average of these A-scans was taken to obtain the final noise-suppressed image and the weights were determined according to the correlation between the corresponding A-scans extracted from the series of B-scans. A small weight was given if the correlation between the correspond-

ing A-scans was poor. The practical challenge of this approach is the very time consuming task of obtaining a number of consistent B-scans using a TD-OCT [28]. Sakamoto et al. [87] utilized a custom software (LabVIEW 7.1, National Instruments, Austin, TX) to average a set of 12 B-scans obtained from an identical retinal location using a SD-OCT imagers. Mayer et al.[88] took the wavelet transform of each B-scan and then calculated the weighted average of wavelet coefficients of multiple frames. The final noise-free image was reconstructed from the averaged wavelet coefficients. They gained a much better SNR ratio by averaging only 8 B-scans rather than simple averaging. In fact, the SNR acquired by averaging 8 B-scans using wavelet technique was comparable to that of simple averaging of 29 B-scans. However, these algorithms require repeated imaging of the same retinal location and therefore are not very efficient in terms of scanning time. Chitchin et al. [89] implemented another wavelet transform method called double-density dual-tree complex wavelet transform to average 7 neighbouring B-scans in order to suppress speckle noise. The final processed image was comparable to the high-SNR averaged image obtained from Spectralis (Heidelberg Engineering, Heidelberg, Germany) OCT system by averaging 50 frames. The Spectralis OCT has a built-in image stabilization and averaging system which can directly produce a high-SNR image by capturing and averaging a series of frames from the same retinal location [90]. Fang et al. [90] exploited sparse representation technique to learn a dictionary from a high-SNR image and used this dictionary to denoise other low SNR images. In another study, Fang et al. [10] utilized structural clustering to learn dictionaries from many clusters of training set high SNR images and then used these learned dictionaries to denoise similar clusters of a low SNR image. They performed sparse coding of 5 neighbouring B-scans and took weighted average to further suppress the speckle noise. The sparse representation technique does not require multiple imaging, and therefore, significantly reduces the image acquisition time. However, the dictionary learning process is tedious and time consuming. The learned dictionary may be less effective if the retinal location and/or pathologies of the eyes are different. In such cases, a new dictionary that better correlates with the anatomical or pathological structure of the test image is required.

4.3 Proposed Multi-frame SD-OCT Denoising Approach

We propose a multiscale denoising algorithm that utilizes both the self-similarity and sparsity for denoising SD-OCT images. Figure 4.1 shows the flowchart of the proposed algorithm. In this approach, multiple B-scans taken from the nearby locations of the SD-OCT volume were used. There is a high degree of correlation among the neighboring B-scans of SD-OCT volume; therefore, these neighboring B-scans can be used to decrease speckle

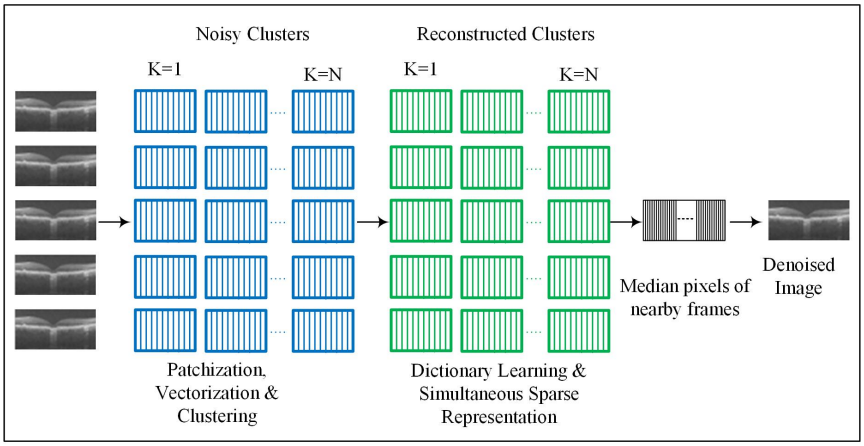


Figure 4.1: Flowchart of the proposed multiscale sparsity-based denoising method.

noise from SD-OCT images. Fang et al. [10] assumed that similar patches from neighboring B-scans can have similar sparse coefficients but with slightly different in magnitudes. We made the same assumption and denoise and combine 5 nearby SD-OCT images to acquire a high SNR B-scan. First, B-scans from very close locations of the SD-OCT volume were extracted and their correlation was examined. If the neighboring patches of the B-scans did not show high degree of correlation, then they were removed.

The test image and its neighboring images were first divided into overlapping patches. Let us consider y_j is the patch created from the test image and $\{y_{j+w}\}_{w=-W}^{W}$ is the corresponding patch from the neighboring slices. These patches were converted to column vectors. Patches from each image were grouped into various structural clusters using k-means clustering method. After structural clustering, a KSVD dictionary was learned from each structural cluster. In case of NL dictionary, a fixed universal dictionary was used for all the clusters. Let D_i be the dictionary created from the test images for i^{th} cluster. The similar patches from the nearby slices were combined into one matrix (one patch from each nearby slice create a matrix that has 5 columns) and simultaneously decompose by solving the following equation

$$\begin{aligned} \{\hat{x}_{j+w}\}_{w=-W}^{W} = & \min_{\{x_{j+w}\}_{w=-W}^{W}} \quad \sum_{w=-W}^{W} \|y_{j+w} - D_i x_{j+w}\|_2 \\ \text{subject to} \quad & \|x_{j+w}\|_0 \leq L,\ w = -W, \cdots, W \end{aligned} \tag{4.15}$$

where x_{j+w} is the non-zero coefficient vector and L is the sparsity constant. The above

problem can be solved using Simultaneous Orthogonal Matching Pursuit (SOMP) developed by Tropp et al. [91]. The SOMP yields sparse representation matrix $\{\hat{x}_{j+w}\}_{w=-W}^{W}$ from the matrix created from the nearby patches in which the indices of non-zero coefficients remain the same while their magnitude varies. The patches were then reconstructed from their sparse vectors.

$$\hat{z}_{j+w} = D_i \hat{x}_{j+w}. \tag{4.16}$$

Finally, median pixels of the reconstructed patches of the nearby frames were chosen which further enhanced the denoising. We selected the median pixels from the nearby frames because this approach is equivalent to the median filtering and median filtering is more robust to an inaccurate image registration as well as to the non-Gaussian noise present in OCT images. The same process was repeated to all the patches in each cluster. In case of KSVD dictionary, a different KSVD dictionary was used for each cluster; however, in case of NL dictionary the same universal dictionar was used for all the clusters. Finally the vectorized form of the reconstructed patches were returned to their original rectangular form. The patches were highly overlapped therefore the overlapping portion of the patches was averaged to reconstruct the image.

4.4 Experimental Results

The proposed denoising algorithm examined two different types of SD-OCT images: synthetic images and real experimental images. The images were provided by Dr. Sinu Farsiu, Duke University, NC, USA [92]. Dr. Farsiu and his group used these images to simultaneously denoise and interpolate SD-OCT images. For detailed information about the image acquisition process we refer the reader to the original paper by Fang et al. [10]. In brief, the synthetic images were generated by subsampling HR images in both random and regular patterns. Thus, the number of A-scans in each B-scan of the SD-OCT image is reduced. The real experimental images were directly taken from human and mouse retina. These images were acquired at relatively low sample rate using a regularly sampled pattern.

The parameters in the proposed algorithm were set by experience. The number of nearby images was set to 4. Two B-scans above and two B-scans below the test B-scans were considered. The patch size was set to 16×4 pixels. The total number of clusters was 90. The NL dictionary was a universal dictionary; therefore the same dictionary was used for all the clusters. In case of adaptive dictionary, we trained KSVD dictionary from each cluster. SOMP was used for simultaneous sparse coding. The number of non-zero coefficients per patch was set to $L = 3$. The standard deviation of the noise in the test SD-OCT image was estimated by employing a MATLAB code written by Fang et al.[10].

Table 4.1: Image quality metrics calculated from the simulated retinal SD-OCT images.

Image(Method)	**CNR**	**MSR**	**PSNR**	**XCOR**	**SSIM**
Test	1.79 ± 0.40	4.12 ± 0.17	17.70	0.9556	0.1092
Average	6.44 ± 4.49	18.67± 1.46	-	-	-
Denoised (NL)	8.57 ±6.86	25.28±3.46	28.39	0.9966	0.6561
Denoised (KSVD)	9.25±7.52	26.71± 2.42	28.54	0.9967	0.6718

4.4.1 Simulated Retinal SD-OCT Images

The simulated datasets were acquired from the subjects with and without nonneovascular AMD. Simulated dataset were obtained by subsampling the previously acquired HR SD-OCT images by 840-nm wavelength SD-OCT imaging system from Bioptigen, Inc. (Durham, NC). The axial resolution of the imaging system was $\sim 4.5\mu m$ per pixel in tissue [10]. A volumetric image was provided by the instrument by comprising 100 B-scans per volume with 1000 A-scans per B-scan. We utilized central foveal B-scans extracted from nearby locations. In order to compare the performance of our denoising approach with the simple frame averaging method, we utilized the average noise-free image (reference image) created by Fang et al. [10]. The reference image was created by registering and averaging 40 azimuthally repeated B-scans using StackReg image registration plug-in for imageJ software (National Institutes of Health, Bethesda, MD, USA) [10]. The size of the input frame was 450×900 pixels. The PSNR between the average image and the test image was 17.70 db, whereas the PSNR between the average image and the denoised image were 28.39 db and 28.54 db using NL and KSVD dictionaries, respectively. The results showed a considerable improvement in PSNR using the proposed method (Table 4.1). The XCORs between the image obtained by the proposed method and the average image were 0.9966 and 0.9967 using NL and KSVD dictionaries, respectively, which were larger than the XCOR computed between the test image and average image (0.9556). The SSIMs increased from 0.1092 (test) to 0.6561 (NL) and 0.6718 (KSVD). There was significant improvement in the quality of the images in terms of CNR and MSR metrics using the proposed method. The processed images obtained from both dictionaries using the proposed method had better CNR and MSR values compared to those obtained from average image (Table 4.1). The KSVD performed slightly better than the NL dictionary. Figure 4.2 shows the test image, average image and image obtained by the proposed method using NL and KSVD dictionaries.

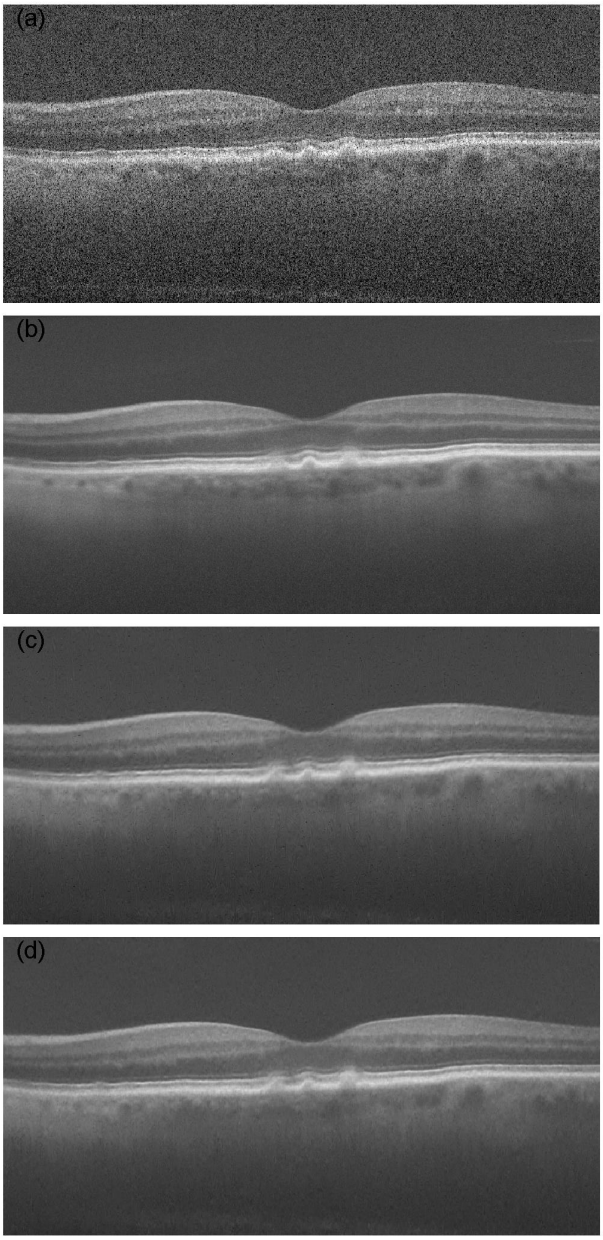


Figure 4.2: Image denoising algorithm applied to simulated SD-OCT images. (a) noisy test image (PSNR=17.70). (b) Average image obtained by registering and averaging 40 azimuthally repeated B-scans using StackReg image registration plug-in for imageJ software. (c) Denoised image by our approach using NL dictionary (PSNR=28.39) and (d) Denoised image by our approach using KSVD dictionary (PSNR=28.54)

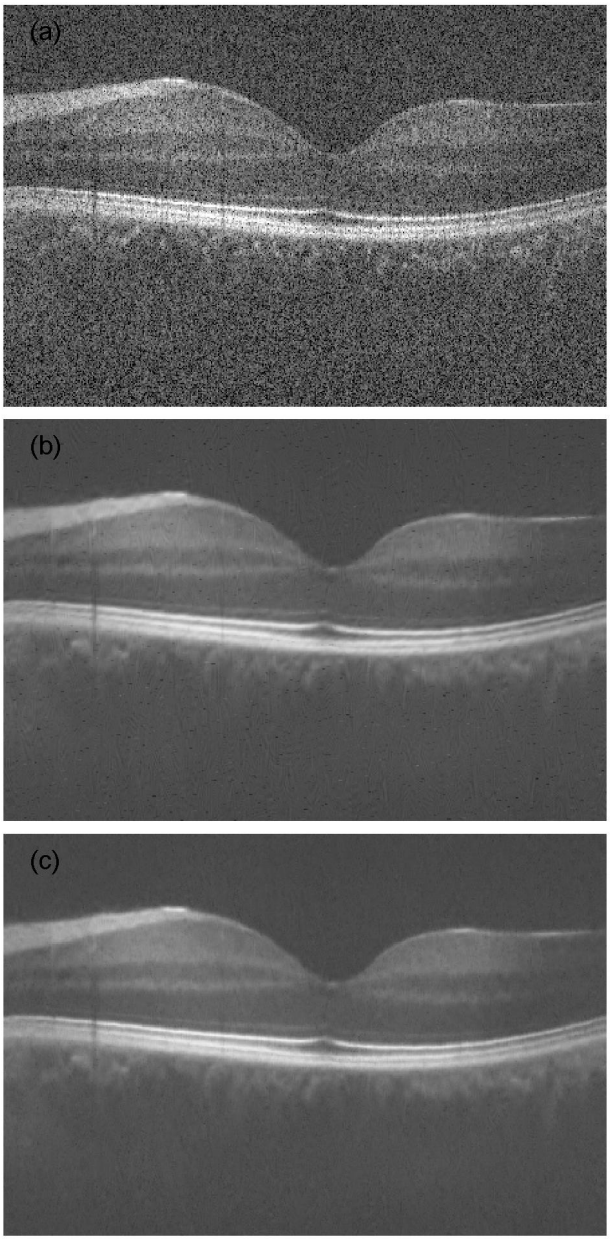


Figure 4.3: Image denoising algorithm applied to human retinal SD-OCT images. (a) noisy test image (CNR=1.43 ±0.50 and MSR=3.25±0.14) (b) Denoised image by our approach using NL dictionary (CNR=3.66 ±1.22 and MSR=12.37±0.81) and (c) Denoised image by our approach using KSVD dictionary (CNR=3.84 ±1.41and MSR=13.12±0.93)

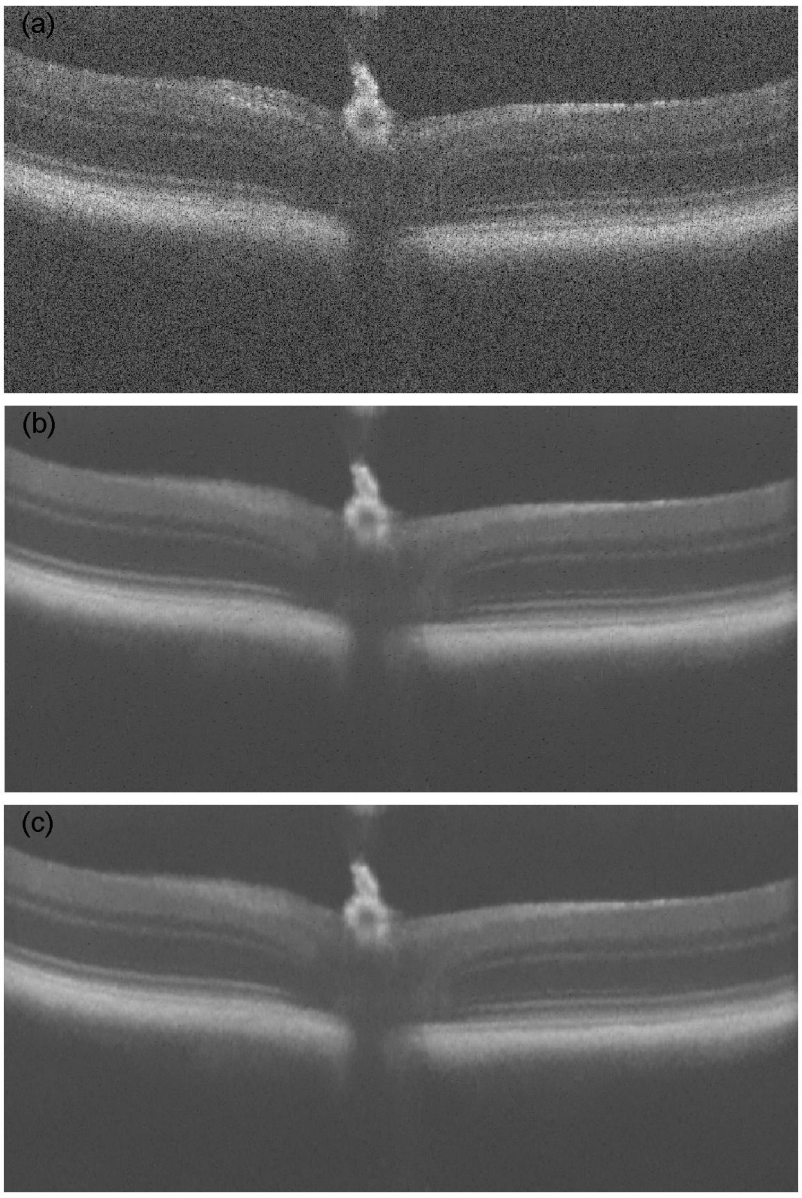


Figure 4.4: Image denoising algorithm applied to mouse retinal SD-OCT images. (a) noisy test image (CNR=1.40 ±0.92 and MSR=3.19±0.24) (b) Denoised image by our approach using NL dictionary (CNR=4.59±1.78 and MSR=15.36±0.64) and (c) Denoised image by our approach using KSVD dictionary (CNR=4.84 ±1.75 and MSR=16.49±0.71)

4.4.2 Human Retinal SD-OCT Images

Human retinal SD-OCT datasets were directly acquired by the Bioptigen SD-OCT imagers described in subsection 4.4.1. The axial resolution was $\sim 4.5\mu m$ and the volumetric image comprised 100 B-scans per volume with 500 A-scans per B-scan [10]. The input frames were extracted from the foveal regions and the size of the frames was 450×450 pixels. The proposed multiscale denoising algorithm was implemented on an experimental dataset of human macula. Figure 4.3 shows the test image and the image obtained by the proposed method using NL and KSVD dictionaries. The visual quality of the image was considerably improved by using the proposed method. Since we didn't have an average image, we were unable to calculate the PSNR, XCOR and SSIM indices. The CNR and MSR metrics obtained by the proposed method were far better than those of the noisy image. This indicates better image quality when using the proposed denoising method. When the performances of the dictionaries were compared, the KSVD dictionaries showed greater improvement in image quality compared to NL dictionary.

4.4.3 Mouse Retinal SD-OCT Images

Mouse retinal SD-OCT datasets were acquired using an ultra-HR SD-OCT system (Bioptigen Envisu R2200) [10]. The axial resolution was $\sim 2\mu m$ in tissue. The input frames were extracted from the optic nerve of a mouse and the size of the frames was 450×1000 pixels. The test image and denoised image from the proposed method using NL and KSVD dictionaries are shown in Figure 4.4. The visual quality of the denoised image obtained from our proposed method was considerably better than that of the noisy image. Similar to the human dataset, we were unable to calculate the PSNR, XCOR and SSIM indices. The CNR and MSR metrics obtained from the proposed method were far better than those of the noisy image. When the performances of the NL and KSVD dictionaries were compared, KSVD showed slightly higher CNR and MSR compared to the NL dictionary.

4.5 Discussion and Conclusion

The proposed multi-frame denoising method works well with both simulated images and real experimental datasets from human and mouse retina. The proposed method significantly increases the value of CNR and MSR metrics. The CNR and MSR values obtained by our approach are better than those obtained by registering and averaging 40 azimuthally

Table 4.2: Image quality metrics (CNR and MSR) calculated from human and mouse retinal SD-OCT images.

	Human Dataset		Mouse Dataset	
Images(Method)	**CNR**	**MSR**	**CNR**	**MSR**
Test	1.43± 0.50	3.25 ± 0.14	1.40 ±0.92	3.19 ±0.24
Denoised (NL)	3.66 ±1.22	12.37 ±0.81	4.59 ±1.78	15.36 ±0.64
Denoised (KSVD)	3.84 ±1.41	13.12 ±0.93	4.84 ±1.75	16.49 ±0.71

repeated B-scans. Regardless of the type of image used (i.e., simulated image or experimental dataset for human and mouse), the proposed method provides high quality noise-free images with minimum artifact. The proposed method also significantly improved the PSNR, XCOR and SSIM indices. The proposed method has many advantages. For example it can be used in the pre-processing stage that will increase the accuracy of other image processing methods such as image segmentation, registration, and interpolation. It does not require multiple scanning of identical locations; rather it reduces noise from the nearby B-scans of SD-OCT volume and takes the weighted average of the denoised B-scans. Therefore, the method significantly reduces the scanning time. The short image acquisition time reduces the patients discomfort and motion artifacts. However, this approach still requires lengthy dictionary learning method if a trained dictionary is used. The computation time was significantly shortened by using NL dictionary; however, the performance was slightly lower. Better denoising might have been possible if the dictionaries were learned from the HR and high SNR training set images.

4.6 Summary

In this chapter, we reviewed various image denoising approaches for SD-OCT images and proposed a new denoising method. The proposed multiscale sparse representation method involves denoising and averaging a series of B-scans extracted from the nearby locations of a 3D SD-OCT volume to obtain a relatively noise-free image. The proposed method provided significant improvement in image quality metrics by reducing speckle noise from the SD-OCT images. In fact, an image obtained by combining 5 nearby B-scans using the proposed method has higher quality metrics than an image obtained by registering and averaging 40 azimuthally repeated B-scans.

Chapter 5

Multi-frame Denoising and Interpolation of SD-OCT Images with Low-Rank Approximations

In chapter 4 we proposed a multi-frame denoising algorithm based on the sparse representation framework. Although the sparse representation technique does not require multiple imaging of the same tissue, and therefore, significantly reduces the image acquisition time, it requires a very tedious and time consuming dictionary learning process. An adaptive dictionary is needed for each cluster if we divide the image into several clusters and incorporate the self-similarity of the image. In addition, if the dictionary is trained for a given image from a set of HR images (such as in [90]), the adaptive dictionary may be less effective for images of other retinal locations and/or pathologies. In such cases, a new dictionary which correlates better with the anatomical or pathological structures of the image is required. In this chapter, we introduce two image restoration algorithms for SD-OCT. The first algorithm is a speckle noise reduction algorithm that utilizes WNNM and non-local self-similarity of images. The second algorithm extends the first algorithm in order to jointly interpolate and denoise SD-OCT images using SVD-based image fusion method.

5.1 Proposed Multi-frame Denoising Approach

As discussed in Chapter 2, a noise-free matrix Z can be recovered from its noisy version Y using the NNM method. That is, the noise-free matrix is obtained by solving the following

energy function [3]

$$\hat{Z} = \min_{Z} \quad \|Y - Z\|_F^2 + \lambda \|Z\|_* \tag{5.1}$$

where $\|Z\|_* = \sum_i \mid \sigma_i(Z) \mid_1$ is the nuclear norm of the matrix Z; $\sigma_i(Z)$ is the i^{th} singular value of Z; $\|.\|_F$ is the Frobenius norm and λ is a positive constant. The solution of Eq. 5.1 can be obtained using soft-thresholding operation on the singular values of noisy matrix [52]. More recent studies have shown that better results are observed with WNNM [51], in which different weights are assigned to the singular values based on their magnitudes. The weighted nuclear norm of the matrix Z is denoted by $\|Z\|_{w,*}$ and is defined as

$$\|Z\|_{w,*} = \sum_i \mid w_i \sigma_i(Z) \mid_1 \tag{5.2}$$

where $w_{i=1}^n \geq 0$ are the non-negative weights assigned to the singular values $\sigma_i(Z)$. In this section, we exploit WNNM method proposed by Gu et al. [51] to reduce speckle noise from SD-OCT images. Figure 5.1 shows a schematic diagram of the proposed approach.

With the developments in technology, SD-OCT is now able to capture 3D dataset by composing a number of B-scans (frames). It has been found that there is a high degree of correlation among neighbouring frames of a 3D SD-OCT volume [10]. Therefore, the similar patches from nearby B-scans can be denoised and averaged to reduce noise. In our denoising method, B-scans from very close locations of the SD-OCT volume were extracted and their correlation was examined. Only B-scans with high degree of correlation were used. We start by extracting overlapping patches from the test frame and its nearby frames. A patch is converted to a column vector by lexicographic ordering. Let y_j^i be the j^{th} patch of the test frame and $\{y_j^{i+b}\}_{b=-B}^{B}$ be the corresponding patches from the nearby frames, where B represents the frames above and below the test frame. A block matching approach is used to find the non-local similar patches across each frame. The non-local similar patches are stacked together to form a matrix Y_k, where $k = 1, 2, \cdots, k$ is the indices of the patch matrices. Each patch matrix Z_k is a low-rank matrix; therefore, it can be estimated from the noisy patch matrix Y_k using low-rank approximation methods [51]. The WNNM method solves this problem by minimizing the following energy function

$$\hat{Z}_k = \min_{Z_k} \quad \frac{1}{\sigma_n^2} \|Y_k - Z_k\|_F^2 + \lambda \|Z_k\|_* \tag{5.3}$$

where σ_n is the standard deviation of the noise in the test image and w is the weight vector. Since larger singular values carry more image information than the smaller ones, the weight vector (w) is determined in a way that the large singular values are shrunk less

than the small singular values

$$w_i = \frac{c\sqrt{n}}{\sigma_i(Z_k) + \epsilon} \tag{5.4}$$

where $\sigma_n(Z_k)$ is the i^{th} singular value of a patch matrix Z_k, n is the number of non-local patches in the matrix, c is a positive constant, and ϵ is a small constant included to avoid dividing by zero [51]. The weight w_i is inversely proportional to the magnitude of the singular value $\sigma_n(Z_k)$. In Eq. 5.4, the singular values $\sigma_n(Z_k)$ are unknown; therefore, the initial singular values $\sigma_n(Z_k)$ are estimated from the singular values of the noisy patch matrix Y_k as

$$\hat{\sigma}_i(Z_k) = \sqrt{\max(\sigma_i^2(Y_k) - n\sigma_n^2, 0)}. \tag{5.5}$$

Equation 5.5 provides the initial estimate of singular values $\sigma_n(Z_k)$. These estimates are used to calculate weights w_i in Eq. 5.4. The weight vector is used to estimate the low-rank matrix Z_k in Eq.5.3. The aforementioned process is applied to each patch of the matrix to get the estimated patch. Finally, a weighted average of the estimated patches $\hat{Z}_j^i$ is taken

$$\hat{z}_j^i = \sum_{b=-B}^{B} a_j^b \hat{z}_j^{i+b} \tag{5.6}$$

where a_j^b is the weight assigned to the j^{th} patch of the input images. The weight is computed by

$$b_j = \frac{\exp\big(-\|z_j^{i+b} - z_j^i\|_2^2\big)/h}{N} \tag{5.7}$$

where N is a normalization factor and h is a predetermined constant. The final estimated patch is reshaped to its original rectangular shape and the overlapping sections are averaged to reconstruct a noise-free image. This method denoises and averages multiple B-scans of 3D SD-OCT volume; therefore, we named this method multi-frame weighted nuclear norm minimization (MWNNM).

5.1.1 Experimental Results

The proposed denoising algorithm was implemented on two different types of SD-OCT images: synthetic images and real experimental images. The SD-OCT images described in Chapter 4 were used to examine the performance of our proposed algorithm. The parameters in the proposed algorithm were set by experience. The number of nearby images was set to 4. Two B-scans above and two B-scans below the test B-scans were

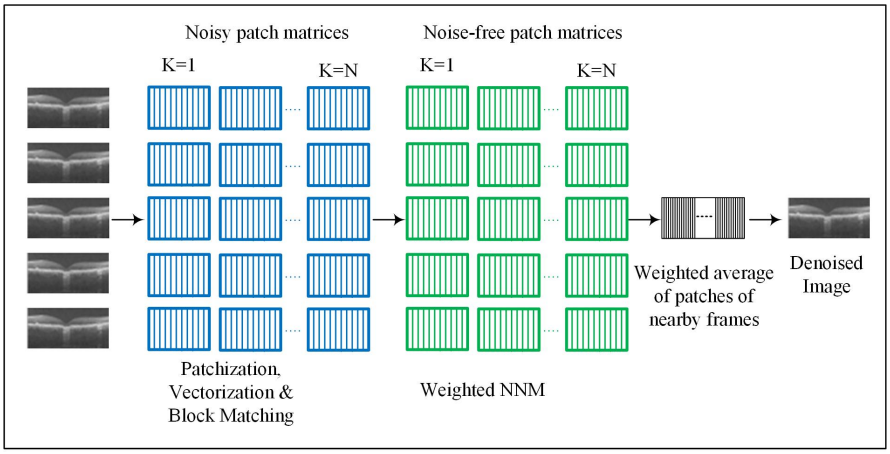


Figure 5.1: A flowchart of the proposed MWNNM method.

considered. The patch size was set to 7×7 pixels. The non-local patch searching window for block matching algorithm was set to 30. The total number of patches in each noisy patch matrix was 70. The parameters c in (5.4) and h in (5.7) were set to $2\sqrt{2}$ and 80, respectively. Iteration number was 8 for simulated dataset and 4 for the real experimental dataset. The final processed patch is the weighted average of five neighbouring image patches in which the WNNM has been performed separately on each patch matrix of individual images. The standard deviation of the noise in the test SD-OCT image was estimated by employing a MATLAB code published by Fang et al. [92]. We also utilized the MATLAB source codes by Zhang [93] for WNNM.

The objective image quality metrics, MSR and CNR were calculated from the test, average and denoised images to compare the performance of the proposed method. MSR was calculated from 4 homogeneous foreground regions (black rectangles in Figure 2.6), whereas CNR was calculated from both the homogeneous and inhomogeneous foreground regions. A total of 8 foreground regions were used to calculate the CNR. XCOR, SSIM and PSNR were also calculated between the reference (average) image and the processed image.

The performance of the proposed method was examined first with the simulated SD-OCT retinal images described in Chapter 4. The simulated SD-OCT images were cropped to 256×512 pixels. The results of the proposed denoising approach implemented on the simulated images are shown in Table 5.1. The image quality metrices, MSR and CNR were calculated from the test, average and denoised images. The CNR and MSR metrics

Table 5.1: Image quality metrics calculated from the simulated retinal images.

Image/Method	**Frames**	**CNR**	**MSR**	**PSNR**	**XCOR**	**SSIM**
Test	5	1.79 ± 0.40	4.12 ± 0.17	17.70	0.9556	0.1114
Average	40	6.44 ± 4.49	18.67± 1.46	-	-	-
Median	5	6.04±4.09	17.55±1.28	27.74	0.9954	0.5848
Mayer et al.	5	9.17±7.55	24.10± 3.95	28.24	0.9965	0.6855
Our Approach	5	9.37±7.78	24.48± 3.31	28.56	0.9966	0.6785
Median	4	5.65±3.67	16.21±1.34	27.24	0.9948	0.5567
Mayer et al.	4	8.41±6.72	21.89± 3.57	27.92	0.9962	0.6690
Our Approach	4	8.74±7.12	22.53± 3.35	28.19	0.9963	0.6647
Median	3	5.05 ±3.06	14.27±0.97	26.26	0.9938	0.5154
Mayer et al.	3	7.23±5.37	18.62± 2.08	27.16	0.9953	0.6372
Our Approach	3	7.71±5.99	19.57± 2.71	27.63	0.9956	0.6439

obtained from our approach were better than those of the average image (Table 5.1). These results show that our approach provides better denoising results by averaging just 5 neighbouring B-scans than the average image which was obtained by averaging 40 images. The PSNR between the average image and the test image was 17.73 db whereas the PSNR between the average image and the image obtained using MWNNM method was 28.56 db (Table 5.1), an indication of a considerable improvement in the PSNR by using MWNNM method. Similarly, the XCOR between the test and average image was 0.9556 whereas XCOR between the average image and MWNNM image was 0.9966 (Table 5.1), suggesting an improved image quality with MWNNM method. The SSIM index improved from 0.114 (between average image and test image) to 0.6785 (between average image and processed image).

The performance of our algorithm was also examined by reducing the number of frames to 4 and 3 and the results are shown in Table 5.1. The results indicate that the image quality metrics decrease with fewer number of frames. The PSNRs between the average image and processed image obtained from 5, 4 and 3 nearby frames were 28.56 db, 28.19 db and 27.63 db, respectively. The XCOR and SSIM indices also declined in a similar fashion. CNR and MSR decreased as the number of frames was reduced. When fewer than 3 frames were used, the CNR and MSR values were lower than those of the average image.

Table 5.2: Image quality metrics (CNR and MSR) calculated from human and mouse retinal images.

		Human Dataset		Mouse Dataset	
Images/Method	**Frames**	**CNR**	**MSR**	**CNR**	**MSR**
Test	5	1.43± 0.50	3.25 ± 0.14	1.40 ±0.92	3.19 ±0.24
Median	5	3.24 ±0.78	10.77 ±1.03	4.08 ±1.86	12.09 ±0.48
Mayer et al.	5	3.74 ±1.32	12.01 ±1.08	5.05 ±1.59	16.38 ±1.20
Our Approach	5	3.87 ±1.53	13.83 ±1.19	4.95 ±1.62	16.77 ±1.74
Median	4	3.15±0.73	10.61 ±1.32	3.87 ±1.80	11.25 ±0.29
Mayer et al.	4	3.67 ±1.27	11.80 ±1.42	4.77 ±1.51	15.15 ±0.48
Our Approach	4	3.88 ±1.54	13.90 ±1.58	4.80 ±1.52	16.20 ±1.78
Median	3	3.00 ±0.58	10.07 ±0.98	3.62 ±1.72	10.39 ±0.66
Mayer et al.	3	3.47 ±1.08	11.08 ±0.89	4.47 ±1.48	13.99 ±1.02
Our Approach	3	3.79 ±1.42	13.49 ±1.38	4.77 ±1.45	16.00 ±1.58

Figure 5.2 shows the test image and three processed images obtained by combining 5, 4, and 3 nearby frames.

We compared the performance of our denoising approach with the wavelet based multi-frame denoising approach proposed by Mayer et al. [88]. We also generated an image by performing median filtering to individual frames (filter size 3×3) and taking a weighted average of the filtered frames. The image quality metrics calculated by these approaches are shown in Table 5.1. The results indicate that among these three methods, median filtering exhibits the worst performance. Our approach showed the largest PSNR, XCOR, CNR and MSR indices whereas the approach proposed by Mayer et al. [88] showed a slightly better SSIM index. The denoising methods were also compared by reducing the number of input frames; however, the performance order remained the same. The average image and the processed images obtained by median filter, Mayer et al. [88] approach and our approach are shown in Figure 5.3.

The proposed approach was also implemented on human SD-OCT retinal images described in Chapter 4. The input frames were extracted from the foveal regions and the frames were cropped to 256×256 pixels. The SD-OCT frames were denoised by median filtering, Mayer et al. [88] multi-frame approach and our approach. Table 5.2 shows the

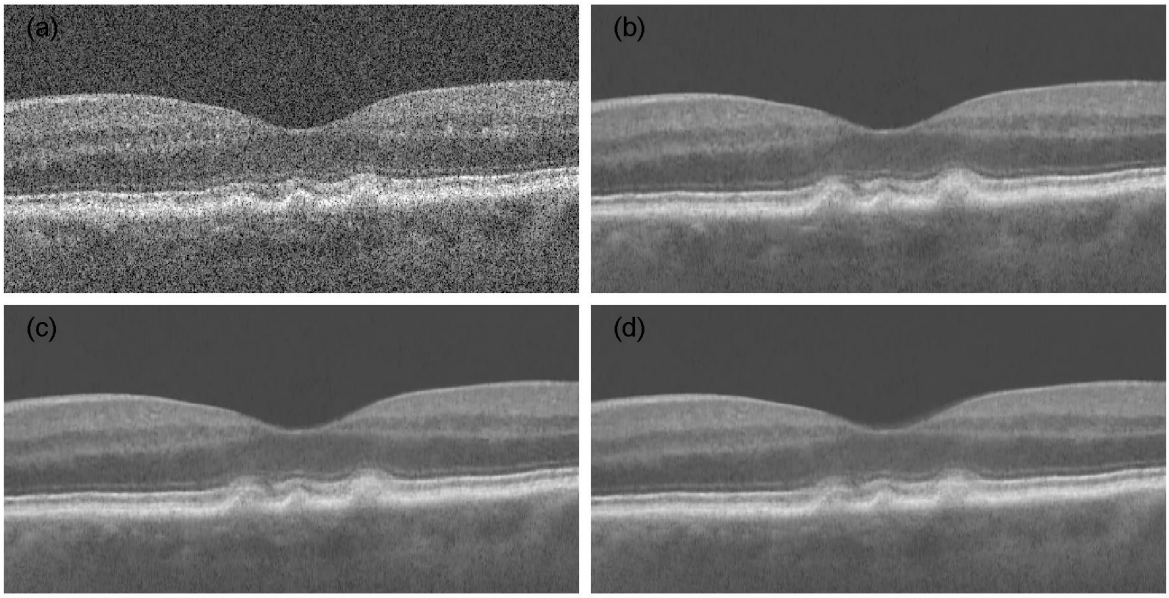


Figure 5.2: Implementation of our denoising approach on simulated SD-OCT images. (a) Test image (PSNR=17.70); images obtained by denoising and averaging (b) 3 frames (PSNR=27.63); (c) 4 frames (PSNR=28.19); and (d) 5 frames (PSNR=28.56) using our approach.

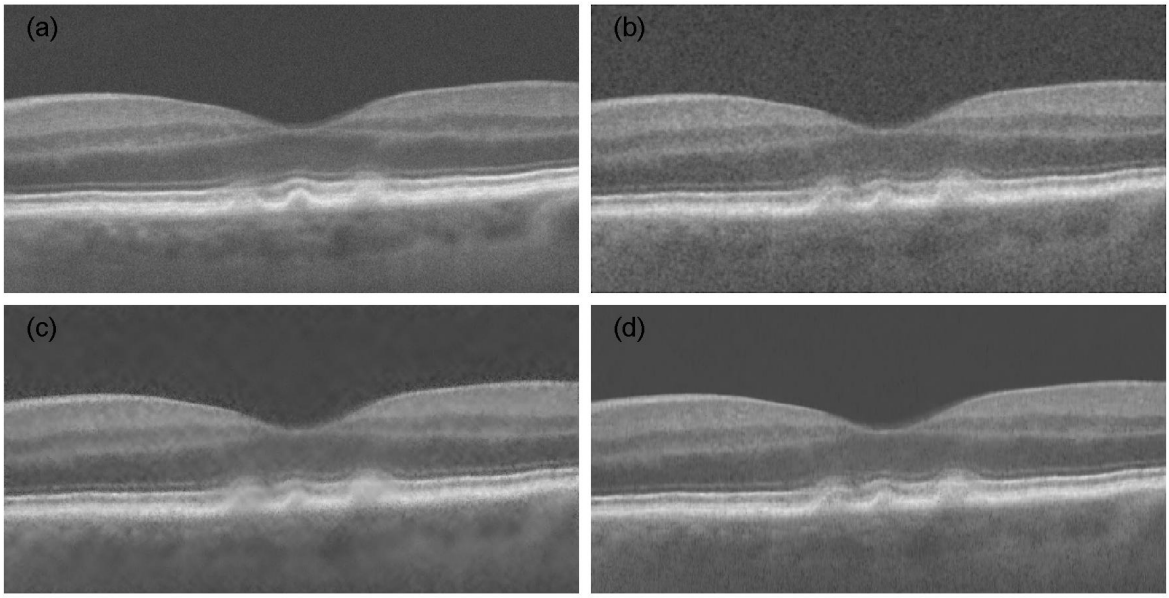


Figure 5.3: Image denoising algorithms applied to simulated retinal SD-OCT images. (a) Reference (average) image. Images obtained by denoising and averaging 5 frames using (b) median filtering and averaging method (PSNR= 26.26); (c) Mayer et al. approach [88] (PSNR=27.16); and (d) our approach (PSNR=28.56).

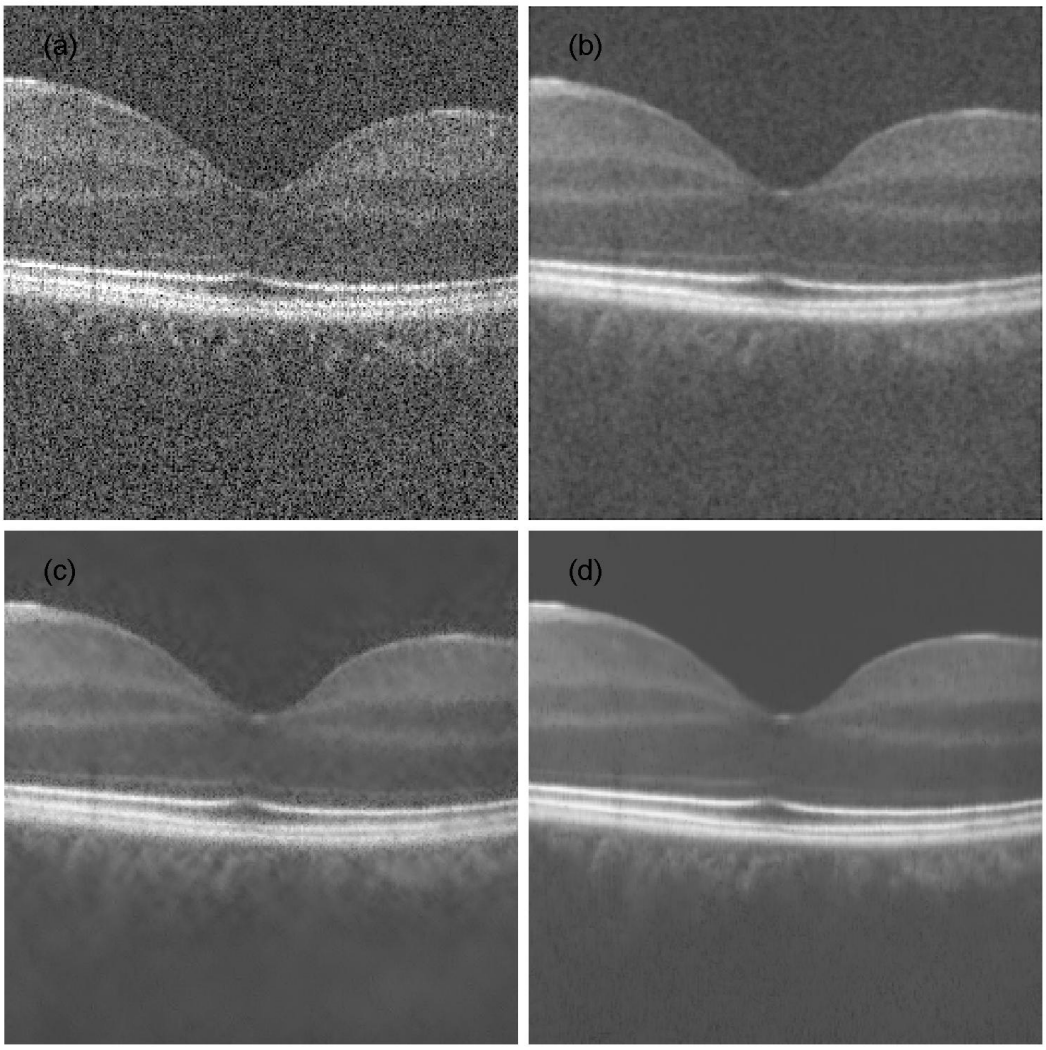


Figure 5.4: Image denoising algorithms applied to human retinal SD-OCT images. (a) Test image (CNR=1.43, MSR=3.25). Images obtained by denoising and averaging 5 frames using (b) median filtering and averaging method (CNR=3.24, MSR=10.77); (c) Mayer et al. approach [88] (CNR=3.74, MSR=12.01); and (d) our approach (CNR=3.87, MSR=13.83).

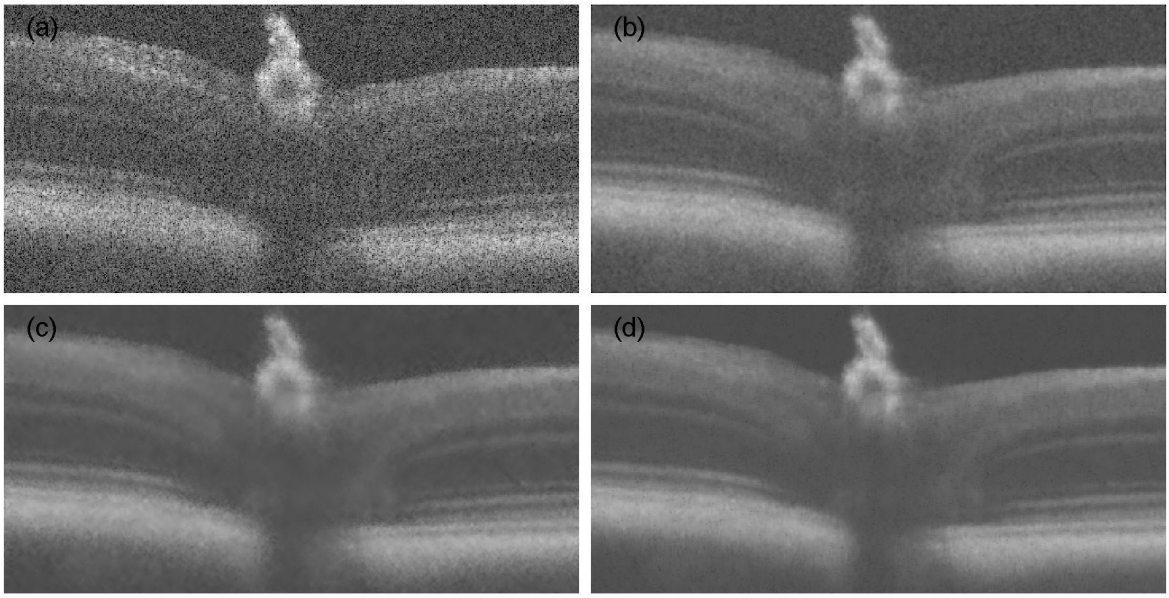


Figure 5.5: Image denoising algorithms applied to mouse retinal SD-OCT images. (a) Test image (CNR=1.40, MSR=3.19). Images obtained by denoising and averaging 5 frames using (b) median filtering and averaging method (CNR=4.08, MSR=12.09); (c) Mayer et al. approach [88] (CNR=5.05, MSR=16.38); and (d) our approach (CNR=4.95, MSR=16.77).

image quality metrics (CNR and MSR) calculated from the denoised images obtained by denoising and averaging 5, 4 and 3 nearby frames. The CNR and MSR metrics obtained by our approach are better than those of the median and Mayer et al. approach. The results showed that CNR and MSR decreased and visual quality of the processed image deteriorated by reducing the number of frames. Figure 5.4 shows the test image and the processed images obtained by denoising and averaging 5 nearby SD-OCT frames using median filtering, Mayer et al. [88] approach and our approach.

Finally, the proposed approach was implemented on Mouse SD-OCT retinal images described in Chapter 4. The input frames were extracted from the optic nerve of a mouse and the frames were cropped to 256×512 pixels. Median filtering, Mayer et al. [88] approach and our approach were implemented on the mouse retinal images. The results are presented on Table 5.2. The test image and the processed image obtained by denoising and averaging 5 nearby frames using the aforementioned approaches are shown in Figure 5.5. The CNR and MSR metrics obtained from our approach are better than those obtained by median filtering or Mayer et al. [88] approach. Mayer et al. [88] showed a slightly better CNR in one condition (i.e., 5 frames); however, our approach showed better CNR for the rest of the conditions and better image quality metrics.

5.1.2 Discussion and Conclusion

In this study, we extended WNNM method in multi-frame to reduce the speckle noise of SD-OCT images. The MWNNM method involves denoising and averaging a series of B-scans extracted from the nearby locations of a SD-OCT volume to obtain a noise-free image. We showed that an image obtained by averaging 5 nearby B-scans with MWNNM method has better image quality metrics than an image obtained by registering and averaging 40 azimuthally repeated B-scans. The CNR and MSR metrics were very small in the test images because of the high level of speckle noise present in the test images; however, when they were denoised and averaged, the image quality metrics significantly increased and exceeded those of an average image obtained by registering and averaging 40 frames. The MWNNM method also significantly improved other image quality metrics such as PSNR, XCOR and SSIM. PSNR, XCOR and SSIM were small when calculated between the test image and the average (reference) image; however, there was a huge improvement in PSNR and SSIM indices when the test images were denoised by MWNNM approach and PSNR and SSIM indices were calculated between the denoised and reference images. We also compared our processed image with the average image generated by median filtering and averaging multiple frames (i.e., weighted average of frames denoised by median filtering). Our method provides significantly better image quality metrics than median filtering.

Similarly, we compared our results with the wavelet based multi-frame denoising approach developed by Mayer et al. [88]. The image quality metrics computed on the denoised image by our approach were greater than those of the Mayer et al. [88] approach in most conditions (Tables 5.1 and 5.2). The image quality metrics computed from denoised image decreased as the frame number was reduced and the level of noise reduction achieved with multi-frame denoising could not be achieved with single frame denoising.

MWNNM works well with both simulated images and real experimental datasets from human and mouse retina. A visual inspection of processed image shows that our multi-frame denoising method significantly reduces the noise without degrading the retinal structures. The retinal structures not very clear in the noisy test scan, are easily identified in the denoised image. For example, the retinal layers not clearly discernible in the noisy B-scan, were clearly detectable in the denoised image (Figure 5.2). Thus the denoised image had greater diagnostic value for retinal pathologies. The denoised image produced by Mayer et al. [88] shows some image artifacts near the edges; however, the denoised image produced by our method is free of such artifacts.

MWNNM is a significantly better method for reducing speckle noise in SD-OCT images than the average image obtained by registering and averaging nearby B-scans. Regardless of the type of image used (i.e., simulated image or real experimental dataset from human and mouse), MWNNM provides high quality noise-free images without generating image artifacts. There are several advantages of MWNNM method. An important advantage is that it does not require any learning steps. Such advantage speeds up the image reconstruction process significantly, and therefore, it can be used in real time. In addition, unlike the sparsity-based method, the MWNNM does not require a high SNR image and works well with images from a lower cost OCT machines, eliminating the scanning time necessary for obtaining a high SNR image. Furthermore, MWNNM does not require multiple scanning of identical locations; rather it takes the weighted average of the noise-free B-scans from the nearby locations. Therefore, the method significantly reduces both the scanning and computation time. The short image acquisition time reduces the patients discomfort and motion artifacts. The MWNNM method can also be used to increase the accuracy of other image processing methods such as image segmentation, registration, and interpolation.

5.2 Proposed Interpolation Approach

As discussed in Chapter 1, since SD-OCT images are corrupted by speckle noise, an efficient image denoising method is required to accurately detect the pathological changes. Besides removing the speckle noise, the resolution of the image should also increase. An HR image

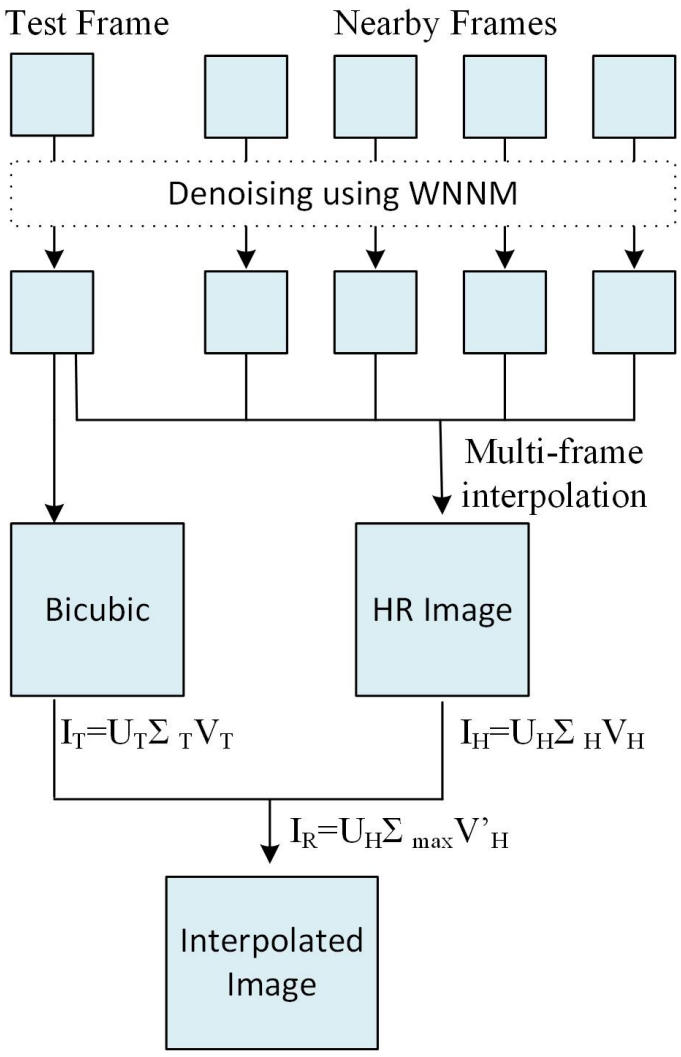


Figure 5.6: A flowchart of the proposed joint denoising and interpolation approach.

has the ability to resolve micron scale retinal structures that allow a better understanding of the anatomy and pathology of the eye [14].

Image interpolation and denoising are two widely used techniques in biomedical image processing. The literature shows that these techniques are mostly used one after another to reduce the computational speed. Some studies [94, 95] first interpolate images and then denoised them, whereas others [96] reverse the order and perform denoising first followed by interpolation. One can also use these methods simultaneously. Recently Fang et al. [10] have proposed a simultaneous image denoising and interpolation technique by exploiting sparse representation technique. They utilized structural clustering to learn dictionaries from different clusters of the training set HR high-SNR images and then used these learned dictionaries to simultaneously denoise and interpolate similar clusters of the LR low SNR images.

In this subsection, we introduce a joint denoising and interpolation algorithm that utilizes multiframe weighted nuclear norm minimization method that was developed in Section 5.1. The proposed joint denoising and interpolation technique consists of two steps. In the first step, multiple frames are denoised, registered and fused together to get a noise-free HR image. The denoising is performed using the WNNM method described in Section 5.1. The denoised frames are registered and fused together using a bicubic interpolation method. Bicubic interpolation method is implemented in this thesis because it is computationally very efficient. A more powerful interpolation method can be adopted to further enhance the performance. In the second step, SVD fusion method is used to exploit the information of HR image to interpolate individual frames. In the SVD fusion method, a block matching approach is used to find similar patches across the image and the similar patches are stacked to form a group. Using block matching the test image and HR image obtained from the first step are divided into several groups. In each group, SVD fusion is applied as follows

$$I_T = U_T \Sigma_T V_T \tag{5.8}$$

$$I_H = U_H \Sigma_H V_H \tag{5.9}$$

where I_T and I_H are the test and HR image. The subscripts T and H represent the test image and the HR image, respectively. The final image is reconstructed using the following equation

$$I_R = U_H \Sigma_{max} V_H' \tag{5.10}$$

where,

$$\Sigma_{max} = \begin{cases} \Sigma_T & \text{if}, \sigma_T > \sigma_H \\ \Sigma_H & \text{if}, \sigma_T < \sigma_H \end{cases} \tag{5.11}$$

where σ_T and σ_H are the singular values of test and HR image, respectively. The approach selects $\Sigma_{max} = \Sigma_T$ if the singular values of test image are larger than that of the HR image, and $\Sigma_{max} = \Sigma_H$ if the singular values of the test image are smaller than that of the HR image. The approach is similar to that applied by Nasir et al. [97]; however, they have not exploited the self-similarity of images. The reconstructed image is an interpolated high SNR image. The flowchart of the proposed joint denoising and interpolation approach is depicted in Figure 5.6.

5.2.1 Experimental Results

The proposed algorithm was examined on simulated SD-OCT retinal images described in Chapter 4. The performance was tested on simulated SD-OCT images from two participants. To differentiate the image sets, they are coded as patient A and patient B images throughout this section. The parameters in the proposed algorithm were kept the same as those in the MWNNM method. The performance of our proposed method was compared with the other SR methods described in Chapter 6; however, the images were corrupted by large amount of noise, and the SR method alone could not provide a high SNR image. There was no improvement in the image quality metrics (such as PSNR, SSIM and XCOR) with the application of SR methods on the SD-OCT dataset. Therefore, the performance of our approach was compared with the state-of-the-art denoising methods plus bicubic interpolation. MSR and CNR were calculated from the images obtained from bicubic interpolation, KSVD [81] + bicubic, BM3D [82] + bicubic, Nonlocally Centralized Sparse Representation (NCSR) [98] + bicubic, single frame WNNM [51] + bicubic and the proposed method. Similar to the MWNNM algorithm, MSR was calculated from the 4 homogeneous foreground regions, and CNR was calculated from 4 homogeneous and 4 inhomogeneous foreground regions. The PSNR, XCOR and SSIM indices were calculated between the processed image and the reference image. The reference image is the average image created by registering and averaging 40 azimuthally repeated B-scans.

Each patient data sets consist of 1 test image and its 4 neighboring images. The images were 450×900 pixels. The images were first subsampled to create LR images of size 225×450 pixels. The LR images were denoised by WNNM method. The denoised images were registered and interpolated using multi-frame bicubic interpolation to obtain an HR image of size of 450×900 pixels. The SVD fusion was performed between the interpolated test image and HR image obtained from the aforementioned process. Table 5.3 provides a quantitative comparison of the proposed method with bicubic interpolation, KSVD [81] + bicubic, BM3D [82] + bicubic, NCSR[98] + bicubic, single frame WNNM [51] + bicubic on patient A retinal images. The results showed that the largest MSR and CNR

Table 5.3: Image quality metrics calculated from the simulated dataset from patient A.

Image/Method	**CNR**	**MSR**	**PSNR**	**XCOR**	**SSIM**
Test	1.72 ± 0.36	3.51 ± 0.20	17.04	0.9446	0.0771
Average	3.99 ± 1.18	11.70± 1.40	-	-	-
bicubic	3.00±0.44	6.86±0.91	21.89	0.9838	0.3390
KSVD [81] ±bicubic	4.09±1.30	9.72± 2.15	23.49	0.9900	0.5942
BM3D [82] ±bicubic	4.10±1.25	9.84± 1.76	23.56	0.9903	0.6114
NCSR [98] ±bicubic	4.14±1.31	9.95± 1.84	23.51	0.9901	0.5880
WNNM[51] ±bicubic	4.21±1.36	9.97± 1.99	23.60	0.9905	0.6168
Our Approach	4.49±1.50	10.92± 1.79	23.94	0.9915	0.6325

values were obtained from the proposed method. Similarly, the largest PSNR, XCOR and SSIM indices between the reference image and the processed image were obtained by the proposed method. Figure 5.7 shows the test image, reference image and processed images obtained from bicubic interpolation, KSVD [81] + bicubic, BM3D [82]+ bicubic, NCSR [98] + bicubic, single frame WNNM [51] + bicubic methods. Similar results were obtained from the patient B retinal images. Table 5.4 shows the MSR and CNR calculated from the processed images and PSNR, XCOR, and SSIM indices calculated between the reference image and the processed image. The proposed approach showed significantly better results compared to the other state-of-the-art methods plus bicubic interpolation. Figure 5.8 shows the test image, reference image and processed images obtained from bicubic interpolation, KSVD [81] + bicubic, BM3D[82] + bicubic, NCSR [98] + bicubic, single frame WNNM[51] + Bicubic methods for patient B retinal images.

5.2.2 Discussion and Conclusion

In this study we showed that a joint denoising and interpolation method is far better for interpolating SD-OCT images. The proposed method involves denoising, registration and fusion of a series of frames extracted from the nearby locations of a SD-OCT volume to create an HR image. The information of the HR image is incorporated to interpolate individual B-scans. This method can be adopted if the images are highly corrupted by speckle noise. The SD-OCT images are examples of such images. Comparing the image obtained from the proposed approach with the directly interpolated one reveals that the joint de-

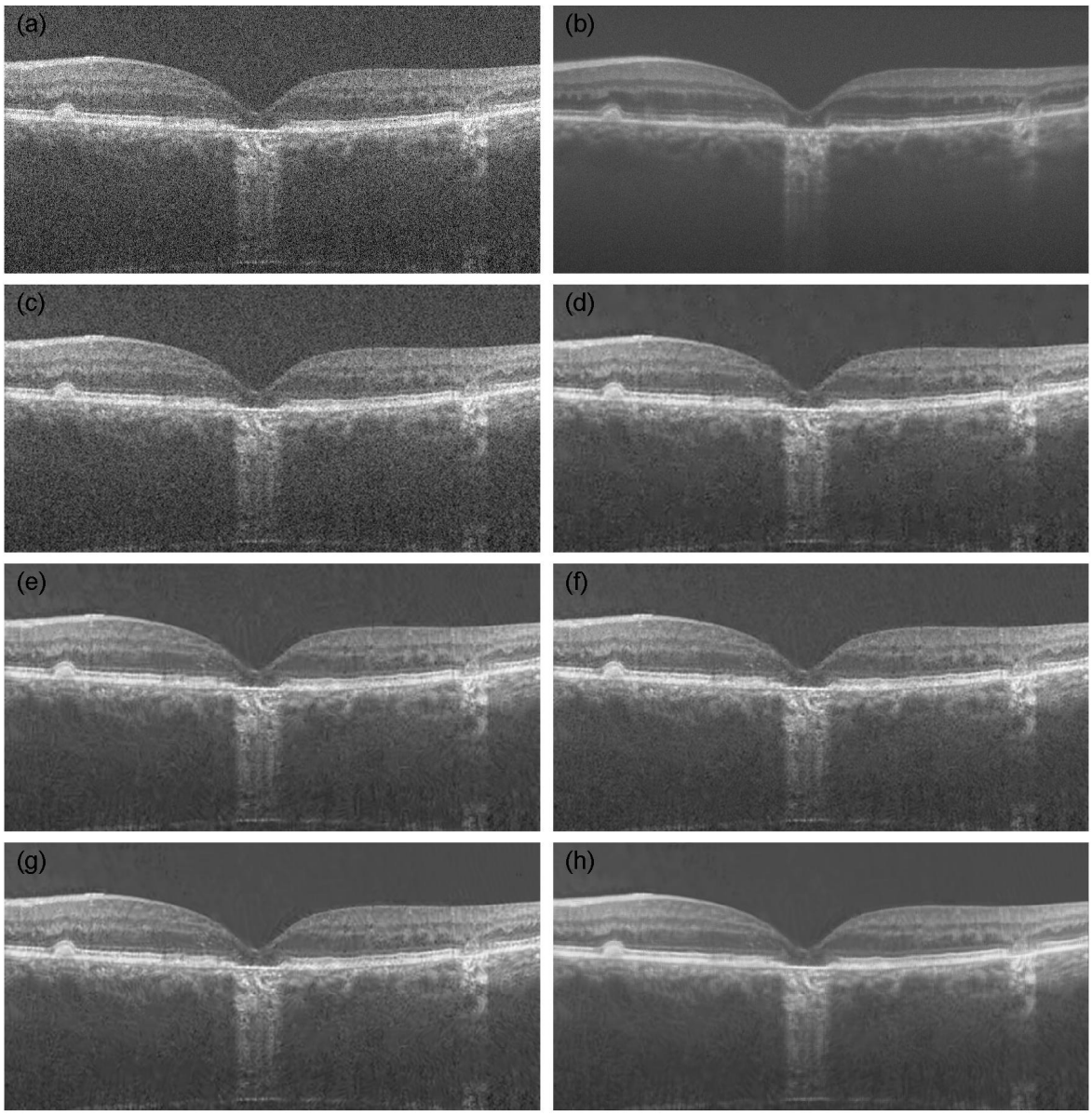


Figure 5.7: Joint denoising and interpolation algorithm applied to patient A simulated SD-OCT images. (a) Noisy test image. (b) Average image. (c) Bicubic reconstruction (PSNR=21.89). (d) KSVD [81] +bicubic reconstruction (PSNR=23.49). (e) BM3D [82] +bicubic reconstruction (PSNR=23.56). (f) NCSR [98]+ bicubic reconstruction (PSNR=23.51). (g) WNNM [51] + bicubic reconstruction (PSNR=23.60). (h) Our approach (PSNR=23.94).

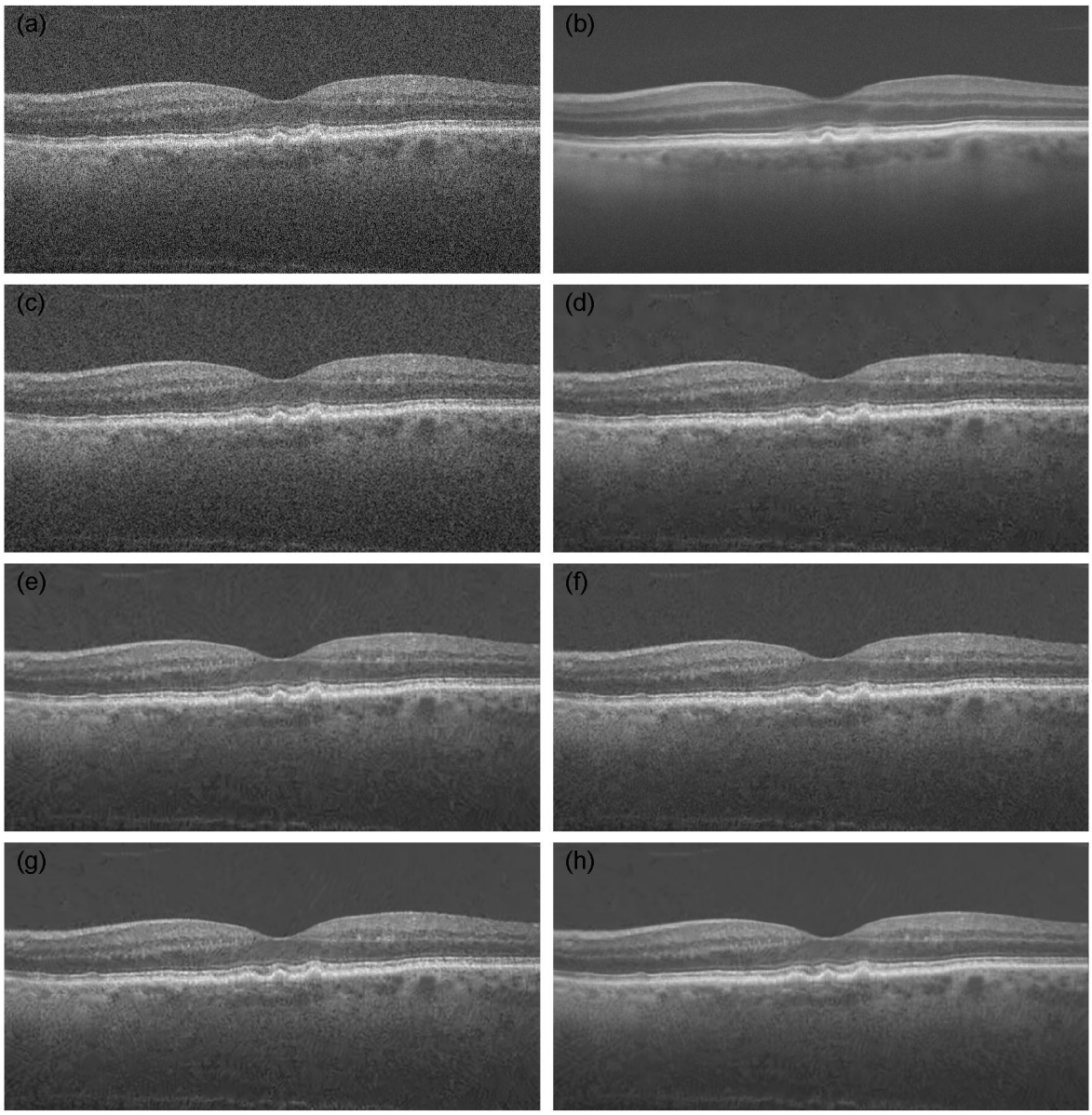


Figure 5.8: Joint denoising and interpolation algorithm applied to patient B simulated SD-OCT images. (a) Noisy test image. (b) Average image. (c) Bicubic reconstruction (PSNR=24.41). (d) KSVD [81] + bicubic reconstruction (PSNR=27.57). (e) BM3D [82] + bicubic reconstruction (PSNR=27.85). (f) NCSR [98] + bicubic reconstruction (PSNR=27.58). (g) WNNM [51]+ bicubic reconstruction (PSNR=27.96). (h) Our approach (PSNR=28.53).

Table 5.4: Image quality metrics calculated from the simulated dataset from patient B.

Image/Method	**CNR**	**MSR**	**PSNR**	**XCOR**	**SSIM**
Test	1.79 ± 0.40	4.12 ± 0.17	17.86	0.9397	0.0867
Average	6.44 ± 4.49	18.67± 1.46	-	-	-
bicubic	3.44±1.60	8.83±0.48	24.41	0.9862	0.3587
KSVD [81] ±bicubic	5.38±3.54	13.19± 1.19	27.57	0.9937	0.6146
BM3D [82] ±bicubic	5.74±3.93	14.55± 1.44	27.85	0.9941	0.6340
NCSR [98] ±bicubic	5.52±3.68	13.52± 1.25	27.58	0.9937	0.6085
WNNM[51] ±bicubic	6.01±4.20	14.69± 1.34	27.96	0.9942	0.6392
Our Approach	6.71±4.90	16.64± 1.45	28.53	0.9950	0.6541

noised and interpolated image has greater diagnostic value for retinal pathologies. The original SD-OCT images are very noisy and the retinal structures are not clearly visible. If we interpolate such images without removing the noise, there will be no advantage in terms of visual quality (see the image interpolated by bicubic interpolation only). However, the retinal structures are easily identified in the interpolated image obtained from joint denoising and interpolation.

The proposed approach significantly increases CNR and MSR metrics. The CNR and MSR metrics obtained by the proposed approach are significantly greater than those obtained by bicubic interpolation, KSVD [81] + bicubic, BM3D [82] + bicubic, NCSR [98] + bicubic, single frame WNNM [51] + bicubic. The proposed method denoised and interpolated the images with minimum image artifacts. We also compared our result with the results obtained by Fang et al.[10] approach. Fang et al. [10] used sparse representation approach to denoise and interpolate the individual B-scans and took the weighted average of similar patches from the neighbouring B-scans. Interpolation is performed by learning a mapping function between the 10 HR training set images and 10 LR training set images. The HR and LR dictionaries and mapping functions are learned from the tranining set images and store. They used different dictionaries for sparse representation and image reconstruction. The LR dictionary is used for sparse representation of test image patch and HR dictionary and mapping function are used for reconstructing the patch from the sparse coefficients. For example let α_l and D_l be the sparse vector and LR dictionary and D_h be the HR dictionary. Let M be the mapping function trained between the LR and HR images. Fang et al. [10] reconstructed the image using the following equation $y = D_h M \alpha_l$.

The PSNRs between the reference image and processed image obtained by Fang et al. [10] were 24.56 db and 29.58 db for patient A and patient B images, respectively, which were greater than that obtained by our method. However, in the method proposed by Fang et al. [10]the processed images were calculated by averaging denoised patches from the 5 nearby images, while in our method the PSNRs were calculated from a single frame. When individual frames were denoised, interpolated and averaged, the PSNR significantly increased and became 24.33 db and 29.15 db for patient A and patient B images, respectively, which were comparable to those of the Fang et al. approach. In addition, our proposed method does not require multiple imaging of a tissue, HR and LR training set images and a tedious and time consuming learning algorithm which significantly increases both the acquisition and computational speed.

5.3 Summary

In this chapter, we proposed two image processing algorithms for SD-OCT images. The first algorithm removes noise from 5 B-scans extracted from the nearby locations of the SD-OCT volume using WNNM method and then averages them to get a noise-free B-scan. The second algorithm jointly denoises and interpolates B-scans from the nearby locations of SD-OCT volume. It consists of two steps. The first step reduces noise from a series of 2D B-scans in order to obtain a relatively noise-free HR image. In the second step the information from the HR image is incorporated to enhance the resolution of individual B-scans using SVD-based fusion approach. Both algorithms showed significant improvement in the image quality in terms of PSNR, SSIM, XCOR, CNR and MSR.

Chapter 6

Comparision of Super-resolution Techniques Applied to Fundus Images

The critical challenge in biomedical imaging is to optimally balance the trade-off between the image resolution, SNR, and the acquisition time. Although acquiring an HR image is possible, it is either expensive or time consuming or both. Resolution is also limited by the physical properties of the imaging device, such as the nature and size of the input source radiation and the optics of the device. SR, an off-line approach for improving the resolution of an image, is free of these trade-offs. Over the past several years several methodologies, such as interpolation, frequency domain, regularization, and learning-based approaches have been developed for SR of natural images. In this chapter, we review some of these methods and demonstrate the positive impact expected from SR of retinal images and investigate the performance of various SR techniques.

6.1 Observation Model

It is important to know the parameters that degrade a retinal image before applying any SR method on it. The blur created either by defocusing camera or motion degrades the retinal image quality. Sampling an object at a frequency less than the highest frequency contained in the object produces aliasing. Also, all retinal images contain some level of noise. These image degradation factors can be incorporated into a mathematical model

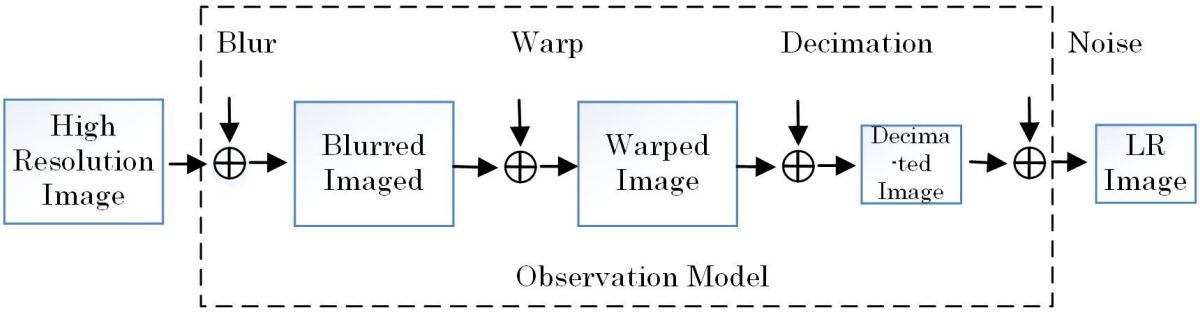


Figure 6.1: Super-resolution observation model. The LR images are the blurred, warped, decimated, and noisy version of an HR image.

that relates the HR image to the observed LR image. To be more precise, let X be an image degraded by motion blur (M) followed by camera blur (B) and decimation effect (D). Furthermore, the image is deteriorated by the white Gaussian noise η created during the acquisition process. The observation model that relates the HR image to the observed LR image is

$$Y^k = DB^kN^kX + \eta^k \tag{6.1}$$

$$Y = HX + \eta. \tag{6.2}$$

The SR methods estimate the image degradation model H and use it to reconstruct an HR image X from a sequence of LR images Y. Figure 6.1 shows a schematic diagram of the observation model. By using different values of k (i.e., different motion parameters, blur, decimation and noise), different LR images can be created from an HR image.

6.2 Multi-frame Super-resolution

In this technique an HR image is reconstructed from a sequence of LR images. A schematic diagram is depicted in Figure 6.2. There are a number of different approaches for reconstructing a single HR image by pooling information from multiple LR images. This thesis includes only the most common reconstruction-based SR approaches.

6.2.1 Interpolation-based Approaches

Interpolation is one of the simplest ways of improving the resolution of an image. It estimates new pixels within an image's given set of pixels. The interpolation has proven

useful in many practical cases. Most of the commercial software, such as Photoshop, Qimage, PhotoZoom Pro, and Genuine Fractals use interpolation methods to resize an image. However, single image interpolation methods are incapable of integrating the information that was lost during the acquisition process [99]. Therefore, we mainly focus on combining a set of LR images to estimate an HR image using non-uniform interpolation methods.

The interpolation-based SR methods involve the following three intermediate steps: registration, interpolation, and restoration. Image registration is the process of geometrically aligning a set of LR images of the same scene with reference to a particular LR image called the reference image. LR images have different sub-pixel displacements and rotations from each other. Therefore it is very important to have an accurate estimation of motion parameters before fusing them to create an HR image. Inaccurate estimate of motion parameters results in various types of visual artifacts that consequently degrade the quality of the reconstructed image. The registration is performed in either the frequency domain or the spatial domain. There are various techniques in each domain for estimating motion parameters. The frequency domain approaches are described in more details in section 6.2.2. As for the spatial domain, Karen et al. [100] proposed an algorithm based on Taylor expansion which estimates the motion parameters with sub-pixel accuracy. Bergen and colleagues [101] proposed a hierarchical framework for the estimation of motion models such as planer and affine methods. Irani et al. [102] developed an interactive multi-resolution approach for estimating motion parameters. To estimate motion parameters, some algorithms map the whole image while others map only the features that are common among the LR images [103]. The HR image and motion parameters can be simultaneously estimated using Bayesian methods. Hardie et al. [104] explain one such approach. The Bayesian approaches are described in more detail in subsection 6.2.3. Besides registration, the interpolation also plays an important role for estimating an HR image. There are many different interpolation methods, yet the complexity of each method depends on the number of adjacent pixels used to estimate the intermediate pixels. Commonly used interpolation methods include nearest neighbor, bilinear and bicubic. Nearest neighbor is the most basic interpolation method that simply selects the closest pixel surrounding the interpolated point. The disadvantage of nearest neighbor is the stair-step shaped linear features visible in the HR image. The bilinear method takes a weighted average of the closest 2×2 neighborhood pixels to estimate the value of the unknown interpolated pixel. Similarly, bicubic takes the closest 4×4 neighborhood pixels to estimate the value of the unknown interpolated pixel. In both of the latter methods higher weights are given to the closer pixels. Since the shifts among the LR images are unequal, non-uniform interpolation methods are required to fuse all LR frames into one HR frame. In 1992 Ur and Gross [105] developed a non-uniform interpolation method for a set of spatially translated LR images using gener-

alized multi-channel sampling theorem [105]. There are many other complex interpolation approaches for resizing a single image, such as Cubic Spline, New Edge-Directed Interpolation (NEDI) [106], and Edge-Guided Interpolation (EGI) [107]. In short, the cubic spline fits a piecewise continuous curve, passing through a number of points. This spline consists of weights and these weights are the coefficients on the cubic polynomials. The essential task of the cubic spline interpolation is to calculate the weights used to interpolate the data. NEDI [106] is a covariance-based adaptive directional interpolation method in which interpolated pixels are estimated from the local covariance coefficients of the LR image based on the geometric duality between the LR covariance and the HR covariance. EGI [107] divides the neighborhood of each pixel into two observation subsets in two orthogonal directions. Each observation subset approximates a missing pixel. The algorithm fuses these two approximate values into a more robust estimate by using linear minimum mean square error estimation. These complex interpolation methods are very efficient and preserve most of the image information; however, their processing time and computational cost is higher than those of the commonly used interpolation methods. The information obtained from registration methods is used to fuse a set of LR images. While fusing the LR frames, pixel averaging methods are used. These methods blur the image; hence, image restoration methods are also needed to remove the blur [99]. Estimation of the blur kernel has an important role in predicting an HR image; however, many SR approaches assume a known blur kernel for simplicity. The known blur kernel can help to estimate an HR image from a set of simulated LR images; however, for real LR images, the motion blur and point spread functions may lead to an unknown blur kernel [108]. Many algorithms are proposed in Bayesian framework to estimate the blur kernel. Recently, Liu and Sun [108] proposed a Bayesian approach of simultaneously predicting motion blur, blur kernel, noise level and HR image. The blind deconvolution algorithm has been used when the information about the blur kernel and the noise level were unknown. The registration, interpolation, and restoration steps in the SR method can be conducted iteratively to achieve an HR image from a sequence of LR images using an Iterative Back Projection (IBP) approach [102]. In this method, an HR image is estimated by iteratively minimizing the error between the simulated LR images and the observed LR images. This approach is very simple and easy to understand; however, it does not provide a unique solution due to the ill-posed inverse problem. Another easily implementable SR approach is the Projection Onto Convex Set (POCS), devised by Stark and Oskoui [109]. In this method, constraints sets are defined to restrict the space of the HR image. The constraint sets are convex and represent certain desirable SR image characteristics, such as smoothness, positivity, bounded energy, reliability, etc. The intersection of these sets represents the space of a permissible solution. Thus, the problem is reduced to finding the intersection of the constraint sets. To find the solution, a projection operator is determined for each convex constraint set. The projection

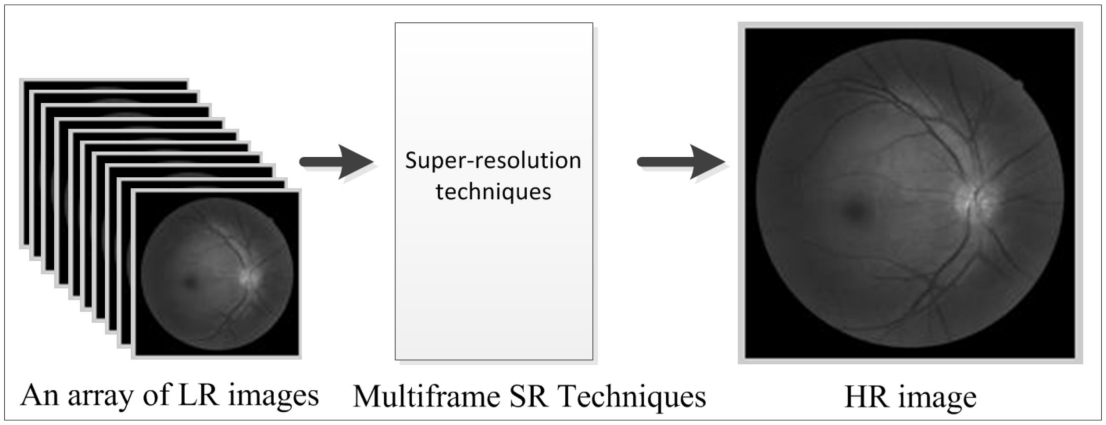


Figure 6.2: The framework of reconstruction-based SR technique. An HR image can be reconstructed by pooling information from many LR images.

operator projects an initial estimate of the HR image onto the associated constraint set. By iteratively performing this approach an appropriate solution can be obtained at the intersection of the k convex constraint sets.

6.2.2 Frequency Domain Approaches

Another popular approach for increasing the resolution of an image is the frequency domain approach. Initially adopted by Tsai and Huang [110], the approach was expanded by many researchers in order to formulate different SR methods. In frequency domain methods, the LR images are first transformed into the Discrete Fourier Transform (DFT). Motion parameters can be estimated in the Fourier domain by measuring the phase shift between the LR images; since spatially shifted images in the Fourier domain differ only by a phase shift [111]. The phase shift between any two images can be obtained from their correlation. Using the phase correlation method, both the planar rotation and the horizontal and vertical shift can be estimated precisely [111]. To minimize errors due to aliasing, only parts of the discrete Fourier coefficients that are free of aliasing are used. After estimating the registration parameters, the LR images are combined according to the relationship between the aliased DFT coefficients of the observed LR images and these of the unknown HR image. The data, after fusion, are transformed back to the spatial domain and reconstructed. The advantage of the frequency domain method is that it is easy and best suited for the aliased images since aliasing is easier to remove in the frequency domain than in the spatial domain.

The disadvantage of the frequency domain method is that the observation model is limited to global motion, so it works only for planar shifts and rotations [111]. The DFT has been replaced by DCT [112] and Discrete Wavelet Transform (DWT) [113] to minimize the reconstruction error.

6.2.3 Regularization-based Approaches

SR is an underdetermined problem with many solutions. Another interesting approach for solving this ill-posed problem is utilizing a regularization term. The regularization approach incorporates the prior knowledge of the unknown HR image to solve the SR problem. Deterministic and stochastic approaches are two different ways of implementing regularization. The deterministic approach introduces a regularization term, which converts the ill-posed problem to a well-posed one

$$X = \min \sum_{k=1}^{N} \|Y^k - H^k X\|^2 + \lambda R \tag{6.3}$$

where R is the regularization term and λ is regularization constant. Various regularization terms have been utilized to solve this ill-posed problem. The constrained least square regularization method uses smoothness, and regularized Tikhonov least-square estimator uses l_2-norm as regularization [114]. The l_2-norm does not guarantee a unique solution. Farsiu et al. [115] exploited an alternative l_1-norm minimization for fast and robust SR. Zomet and colleagues [116] described a robust SR method for considering the outliers. Recently, Mallate and Yu [117] proposed a regularization-based SR method which uses adaptive estimators obtained by mixing a family of linear inverse estimators.

The stochastic approach, especially Maximum A-Posteriori (MAP) approach, is popular since it provides a flexible and convenient way to include an a priori and builds a strong relationship between the LR images and the unknown HR image. The method proposes to find the MAP estimation of the HR image X^{MAP} for which a-posteriori probability $P(X \mid Y)$ is a maximum [99]

$$\hat{X}^{MAP} = \min_{X} \quad P(X \mid Y) \tag{6.4}$$

Using Bayes theorem, the above equation can be written as [104]

$$\hat{X}^{MAP} = \min_{x} \quad [\log P(Y \mid X) + \log (P(X))] \tag{6.5}$$

where $P(Y \mid X)$ is the likelihood function and $P(X)$ is the priori. Markov Random Field (MRF) is commonly used as the prior model and the Probability Density Function (pdf) of noise is calculated to determine the likelihood function [118]. The HR image is computed by solving the optimization problem defined in Eq. 6.4. Several models such as Total Variation (TV) norm, [119], l_1-norm [120] of horizontal and vertical gradients, Simultaneous Auto Regressive (SAR) norm [121] have been used for the prior image model. Similarly, other frequently used prior image models are Gaussian MRF model, Huber MRF model, discontinuity adaptive MRF model, the two-level Gaussian non-stationary model, and Conditional Random Field (CRF) model [99].

All the above SR techniques assume that the blurring function is known. The blur can be modeled by convolving the image with the point spread function; however, it requires a strong prior knowledge of the image and blur size. The blind deconvolution algorithm can be used when the information about the blur and the noise are unknown. The blind deconvolution SR methods recover the blurring function from the degraded LR images and estimate the HR image without any prior knowledge of blur and the original image [122, 123]. The blur is calculated from another regularization term as shown in the following equation

$$E(X,h) = \sum_{k=1}^{k} \|DH_kX - Y_k\|^2 + \alpha Q(X) + \beta R(h). \tag{6.6}$$

The first term is the fidelity term, and the remaining two are regularization terms. The regularization $Q(X)$ is a smoothing term while $R(h)$ is a probability density function (pdf) term. The regularization is carried out in both the image and blur domain.

6.3 Single-frame Super-resolution

Learning-based SR methods extract the high frequency information, (lost during image acquisition process), from the external sources (training set images) and integrate this information with the input LR image to acquire a super-resolved image [124, 125, 126] . Figure 6.3 shows the schematic diagram of learning-based SR. The training set consists of many HR images and their simulated LR versions. The performance of the learning-based SR methods highly dependent on the training set data. The training set images should have high frequency information, and be similar to the input LR image [124]. To reduce the computational complexity, the learning-based methods are usually performed on the image patches. The learning-based SR methods include the following four stages as depicted in Figure 6.4. First, the HR patches and their simulated LR version are stored in the training

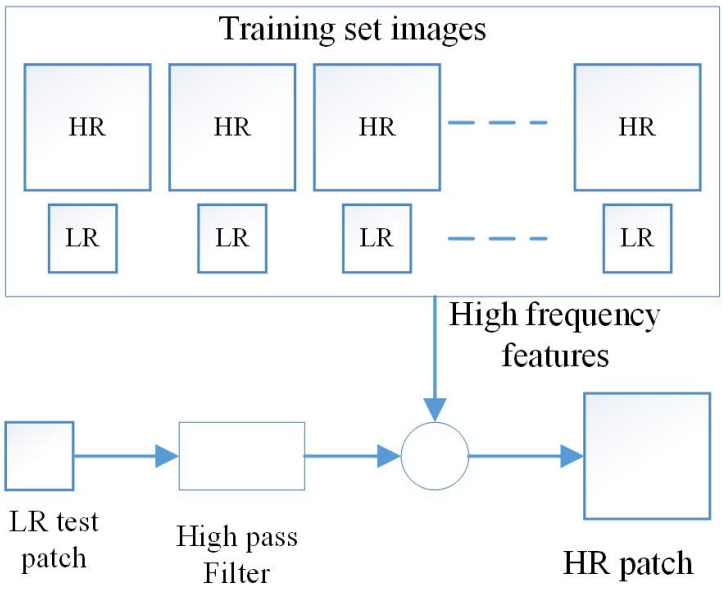


Figure 6.3: An illustration of learning-based SR.

set in pairs. The features of the training set patches are extracted. A number of different types of feature extraction models can be used such as luminance values, DCT coefficients, wavelet coefficients, contourlet coefficients, PCA coefficients, gradient derivatives, Gaussian derivatives, Laplacian pyramid, steerable pyramid, feature extracted from bandpass filter, low and high pass frequency components, quaternion transformation, histogram of oriented gradients, etc. A summary of various feature extraction models is found in [124]. Second, features of the input LR patches are extracted. Third, the features extracted from the input patches and training set patches are matched and the best matched pair from the training set is selected. In recent years a number of learning methods for matching the features have been proposed. The most common learning models are Best Matching, MRF, Neighbor Embedding, and Sparse Representation Model [124]. Fourth, the learned HR features are integrated into the input LR patch to achieve a super-resolved patch. Finally, all super-resolved patches are combined to generate the HR image.

The example-based (EB) SR method proposed by Kim and Kwon [126] has outperformed several state-of-the-art algorithms in single image SR. This method is based on the framework of Freeman et al. [125] which collects pairs of LR and HR image patches in the training stage. In the learning stage, each LR patch of the input image is compared to the

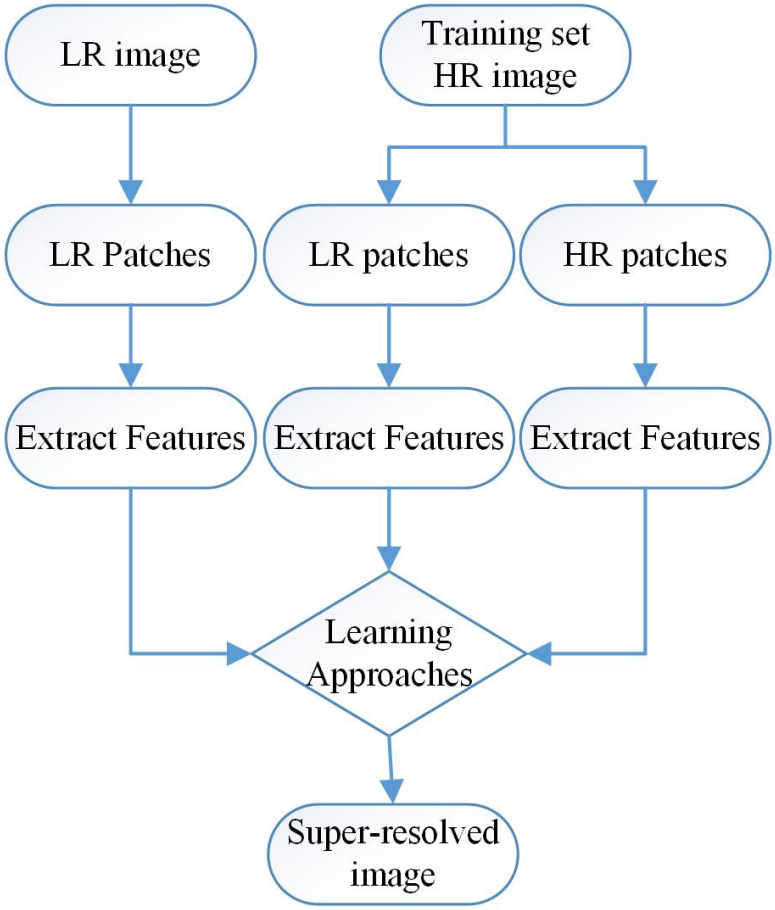


Figure 6.4: Flow chart of learning-based SR.

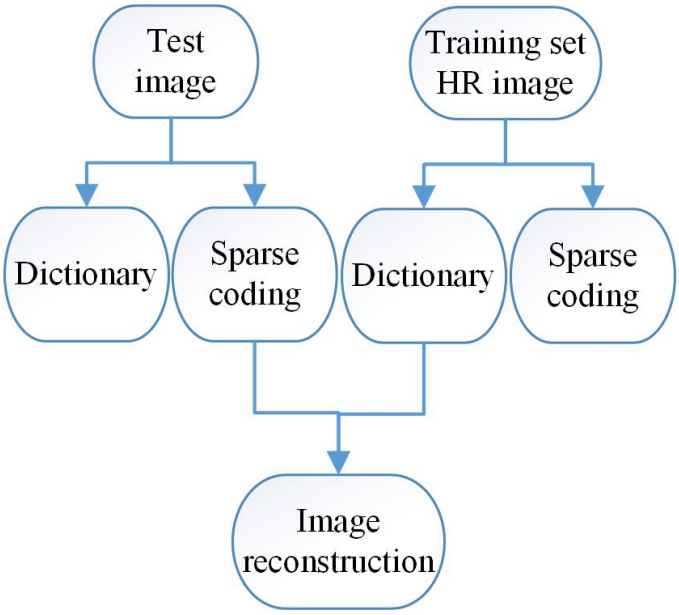


Figure 6.5: Flow chart of SR via sparse representation algorithm.

stored training set LR patches, and using a nearest neighbor search method a nearest LR patch and its corresponding HR pair are selected. However, this approach often results in a blurred image. Kim and Kwon [126] modified this approach by replacing nearest neighbour search with sparse kernel ridge regression. In their approach, kernel ridge regression is adopted to learn a map from input LR patch to training sets HR and LR patch pairs. This method also produces some blurring and ringing effects near the edges which can be removed using post processing techniques [126].

Sparse representation has become a major field of research in signal processing. Utilizing this approach, several researchers have proposed learning-based SR algorithms [127, 128, 80, 129, 84, 98]. Sparse representation based SR computes the sparse approximation of input LR patch and uses the coefficients of approximation to estimate an HR patch. In this method, two dictionaries D_h and D_l are jointly trained from HR and LR patches. There is a need to enforce the similarity of sparse coding between the LR ($j = D_l\beta$) and HR patch ($l = D_h\alpha$). The dictionary extracted from the HR patch D_h is applied with the sparse representation of the LR patch ($D_h\beta$) to recover the super-resolved patch. The schematic illustration of SR via sparse representation as proposed by Yang, Wright, Huang, and Ma [127] is shown in Figure 6.5.

In sparse representation-based approach, the final super-resolved image patch is generated from the combination of sparse coefficients of the LR patch and the HR dictionary; the performance of the method depends upon both the sparse coefficients of LR patch and the HR dictionary. Many researchers have proposed new algorithms to better estimate the HR dictionary and sparse coefficients of the LR image. Dong et al. [80] proposed a clustered based sparse representation model called Adaptive Sparse Domain Selection (ASDS) to improve the dictionary. In this approach, the image patches are gathered into many clusters and a compact subdictionary is learned for each cluster. For each image patch, the best subdictionary can be selected that can reconstruct an image more accurately than a universal dictionary. In another study, Dong et al. [129] proposed sparse representation-based image interpolation through incorporating the image non-local self-similarities to the sparse representation model. The term self-similarity refers to the similarity of image pixel values or structure at different parts of the image. The algorithm included non-local autoregressive model as a new fidelity term to the sparse representation model which reduces the coherence between the dictionaries, and consequently makes the sparse representation model more effective. Dong and colleagues not only estimated better HR dictionary for each image patch, but they also utilized the image non-local self-similarity to obtain good estimation of the sparse representation coefficients of LR image. Recently, they have proposed two models for extracting sparse coding coefficients from the LR image as close to the original image as possible using non-local sparsity constraints. These are the Centralized Sparse Representation (CSR) model [84] and the Nonlocally Centralized Sparse Representation model (NCSR) model [98].

6.4 Experimental Results

MATLAB software (version R2008a) was used to code and/or to run the programs. The MATLAB codes were downloaded from the websites of respective authors, and the parameters of each method were set according to the values given in their corresponding papers. A computer with the operating system 64 bit version of Windows 7, Intel (R) Pentium (R) CPU G620T 2.2 GHz processor, and 4GB RAM was used to run the simulations. The screen resolution was 1920×1080. Retinal fundus images taken with a fundus camera (Non-Mydriatic Auto Fundus Camera, Nidek AFC-230, Japan) were used to run the simulations. The images were taken from the author who has no ocular pathology. SR approaches were applied separately for real LR images and simulated LR images. Simulated LR images are viewed as the shifted, rotated and downsampled version of an HR image. We cropped 3 important sections (256×256 pixels) from 3 different fundus images to run SR in different

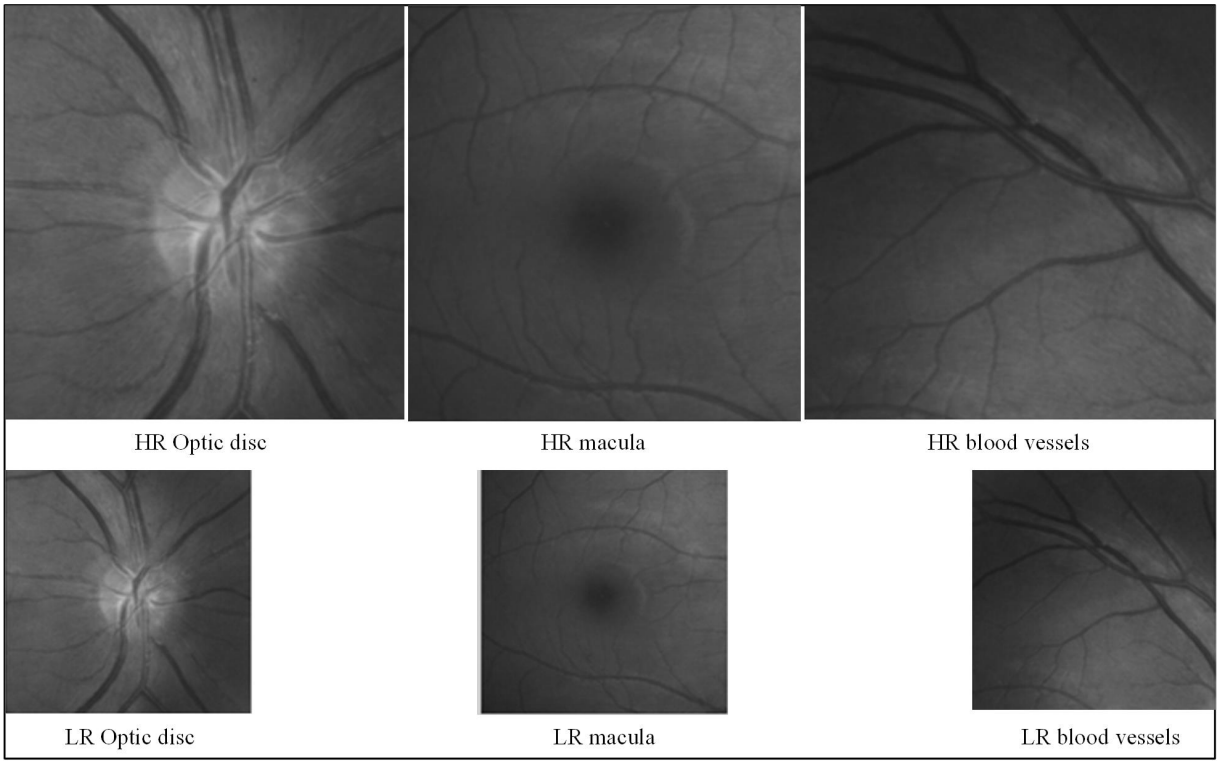


Figure 6.6: HR test images created by cropping three different sections of different fundus images and their corresponding LR versions.

Table 6.1: PSNR and SSIM indices between the original and the super-resolved images obtained from a set of LR simulated images using different multi-frame SR approaches.

	Optic disc		Macula		Blood vessels	
Methods	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Fourier	36.67	0.8944	42.47	0.9596	42.14	0.9579
Robust regularization	39.22	0.9393	44.54	0.9727	43.69	0.9703
IBP	39.79	0.9455	44.51	0.9712	42.90	0.9635
POCS	39.99	0.9481	44.13	0.9681	41.29	0.9517
TV norm	42.57	0.9626	46.13	0.9789	45.33	0.9762
L_1 norm	42.31	0.9606	45.98	0.9781	45.20	0.9757
SAR norm	42.48	0.9632	46.23	0.9795	45.51	0.9777

parts of the fundus image. The cropped sections were the optic disc, macula and retinal blood vessels (Figure 6.6). Four 128×128 pixels LR images were created from these HR images. The shift and rotation parameters were adjusted manually. Horizontal and vertical displacements were set to $(0, 0)$, $(-3, 2)$, $(2, 1)$ and $(-2, 3)$ and rotation angles were set to $(0, 5, 3, -2)$ degrees, respectively, to create 4 LR images. The downsample factor was set to 2. Finally, Gaussian noise was added to the LR images to maintain the SNR of 40dB. The first LR image is the reference LR image which is a downsampled version of the HR image with the shift and rotation parameters at zero. The reference LR images are shown in Figure 6.6 with their original. Figure 6.7 shows all 4 LR images that were created from the cropped optic disc image. We used these simulated LR images to recover the original HR image (resolution 256×256) using various SR methods.

Frequency domain SR approaches [111] were first examined on the simulated LR fundus images. These images were transformed into the Fourier domain, and shift and rotation parameters between the LR and reference images were calculated based on their low-frequency, aliasing part. Shifts were estimated from the central low frequency components in which the number of low frequency components used was 10 and the rotations were estimated from a disc of radius 0.8. By incorporating these motion parameters on the simulated LR images, an HR image was reconstructed using cubic interpolation. Besides cubic interpolation, the performances of IBP [102], Robust Regularization [116], and POCS [109] were also examined in Fourier domain. We employed MATLAB software prepared by Vandewalle, Su and Vetterli [111] to implement these algorithms. For IBP, an upsampled

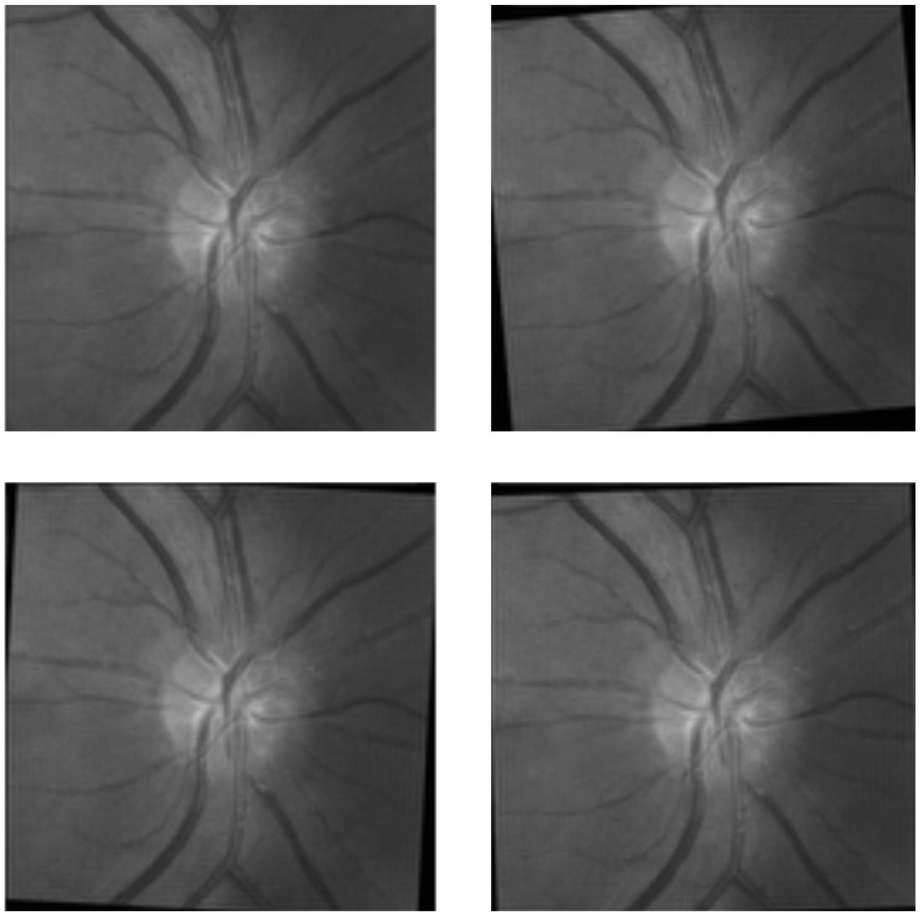


Figure 6.7: LR images (resolution 128×128) obtained from an optic disc using observation model. These images were used to run all the multi-frame SR techniques.

version of the reference LR image was used as an initial estimate of HR image. The upsampling was performed using bicubic interpolation. The IBP created a set of LR images from the initial estimate of HR image using the motion parameters estimated in Fourier domain. The estimate was then updated by iteratively minimizing the error between the simulated LR images and test LR images based on the algorithm developed in [102]. Robust regularization further incorporates a median estimator in the iterative process to achieve better results. We implemented the robust regularization algorithm proposed by Zomet and colleagues [116]. The POCS algorithm which reconstructs an HR image using projection on convex sets was examined only for the planer shift.

Similarly, Bayesian SR methods were studied and their robustness on the LR fundus images was tested for various prior models. We used algorithms and MATLAB software prepared by Villena, Vega, Babacan, Molina and Katsaggelos [121] for the simulation. Total Variation (TV) [119], l_1 norm of the horizontal and vertical gradients [120], and Simultaneous Auto Regressive (SAR) [121] were used as image prior models. The motion parameters and downsampled factor were kept unchanged between the Fourier domain methods and Bayesian methods for fair comparison except for POCS in which planar rotation was not applied. The simulated 4 LR images were used as input. The algorithm utilized hierarchical Bayesian model where the model parameters, registration parameters and HR image were estimated simultaneously from the LR images. Variational approximation was applied to estimate the posterior distributions of the unknowns. The algorithm terminated when either a maximum number of iterations (k=100) or the criterion $\|x^k - x^{k-1}\|^2 / \|x^{k-1}\|^2 < 10^{-4}$ where x^k is the k^{th} estimated HR image, was satisfied. The Bayesian methods showed the highest PSNR value compared to the other multi-frame SR methods. However, the TV norm, l_1 norm of the horizontal and vertical gradients and SAR norm priors model led to over-smooth non-edge regions of the image. Figures 6.8, 6.9 and 6.10 show the results of various multi-frame SR approaches applied to the LR images of optic disc, macula and retinal blood vessels.

Single image interpolation methods were also studied on retinal images. The input LR image was created by direct subsampling of the original image by a factor of 2. The LR retinal image was upscaled to double its size 256×256 using nearest neighbor, bilinear and bicubic interpolations. The interpolated images were compared with the original image. The PSNR and SSIM indices for bicubic method were greater than those of the nearest neighbor and bilinear interpolation. The complex interpolation methods, such as Cubic Spline [117], NEDI [106], and EGI [107] were also applied to the downsampled LR images of optic disc, macula and retinal blood vessels. Noise was not added to the LR image of single image interpolation methods so they showed better PSNR and SSIM indices. The comparisons between various single image interpolation approaches in terms

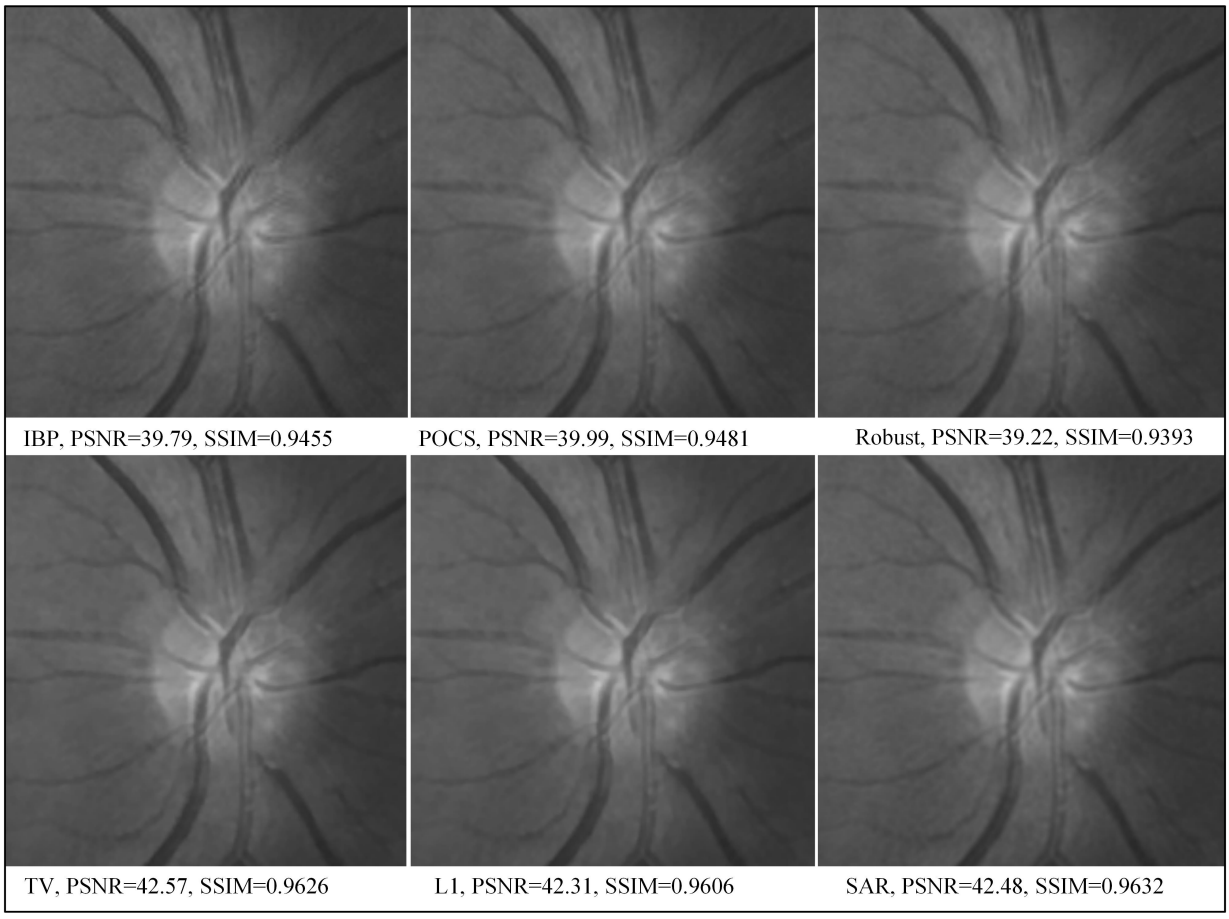


Figure 6.8: Results from different multi-frame SR approaches for LR optic disc images.

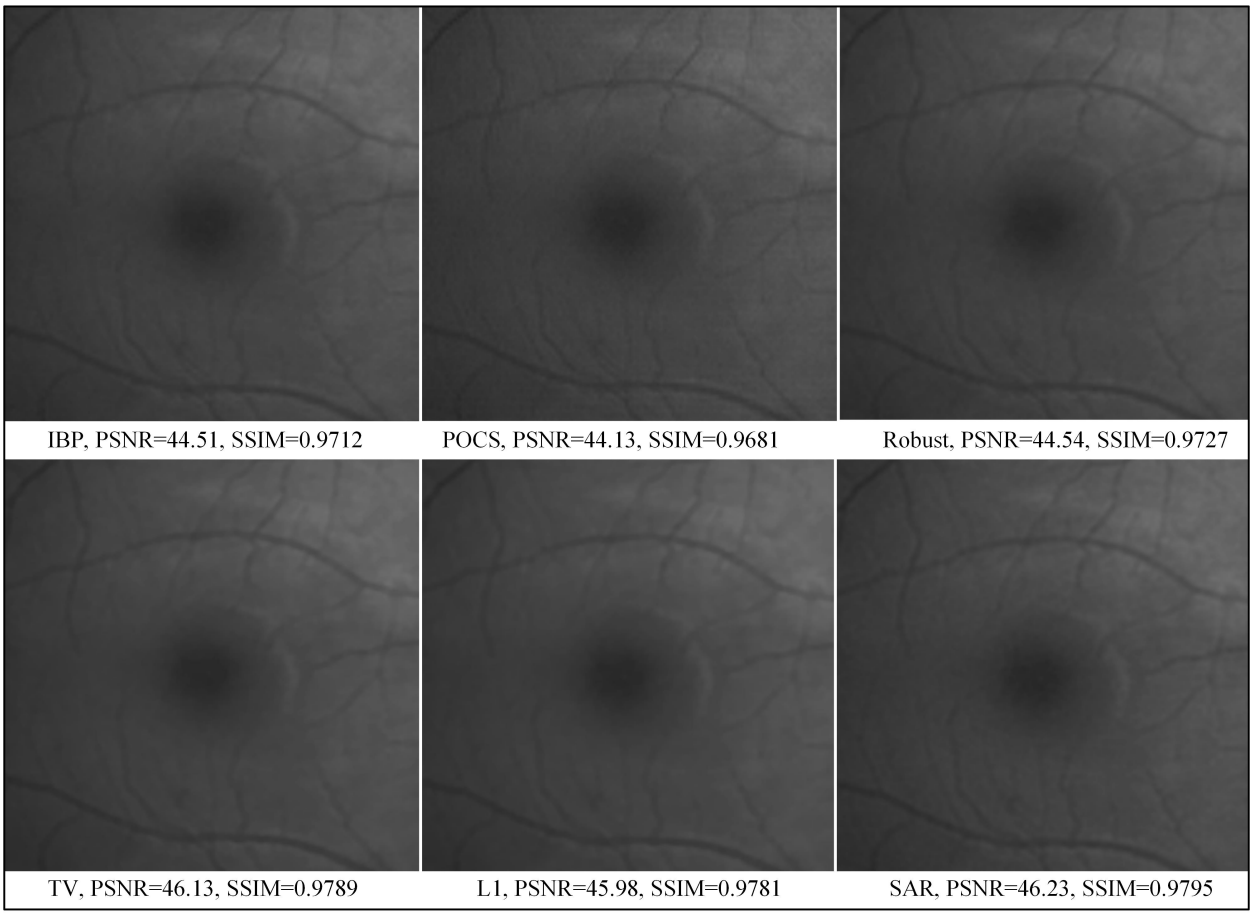


Figure 6.9: Results from different multi-scale SR approaches for LR blood vessels.

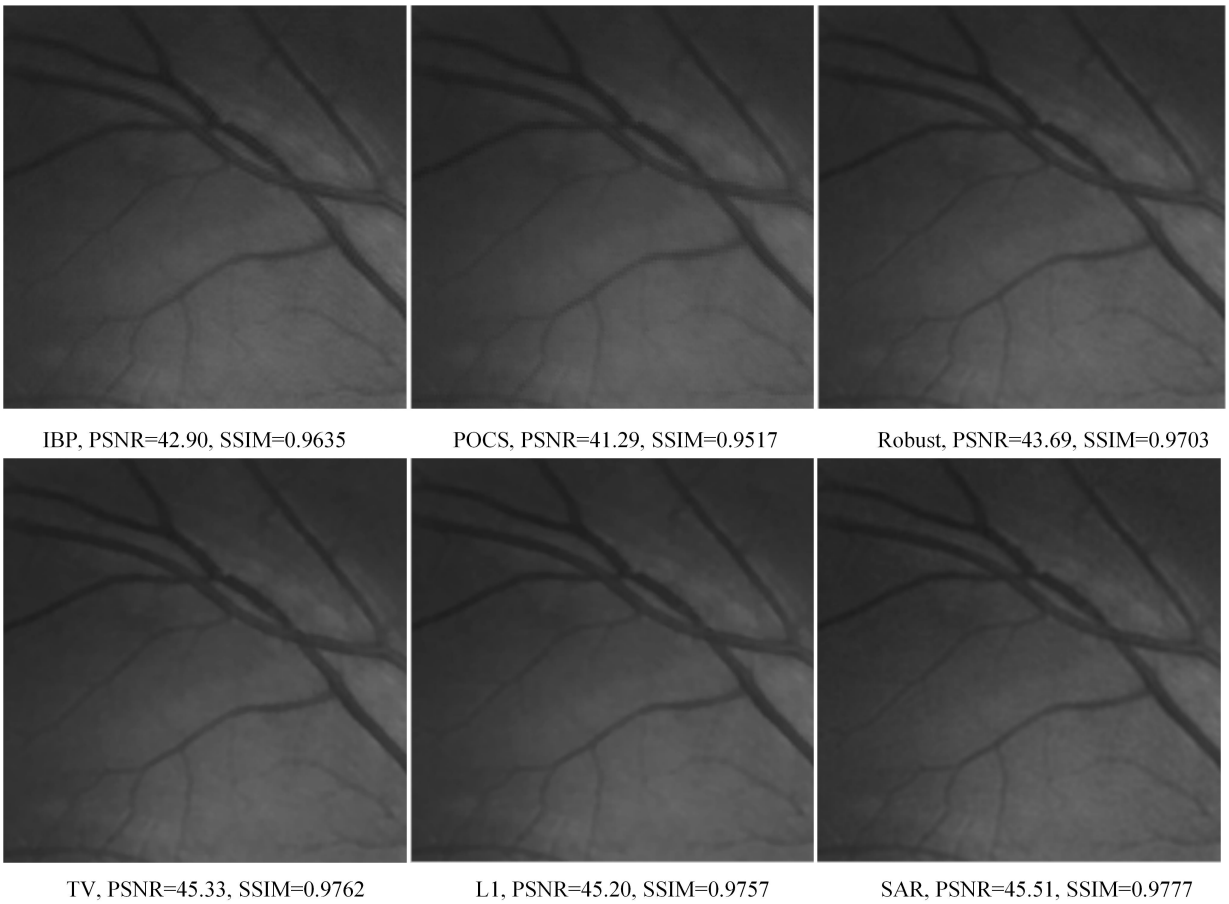


Figure 6.10: Results from different multi-frame approaches for LR macula images.

Table 6.2: PSNR and SSIM indices between the original and the reconstructed images obtained from various single image SR approaches.

	Optic disc		Macula		Blood vessels	
Methods	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
Bicubic	41.87	0.9592	47.06	0.9831	46.01	0.9802
EGI[107]	41.84	0.9593	46.42	0.9814	45.83	0.9797
NEDI[106]	41.88	0.9609	46.50	0.9817	45.83	0.9798
SME[117]	42.71	0.9663	47.01	0.9835	46.64	0.9828
Cubic interpolation	42.14	0.9621	46.60	0.9823	46.20	0.9812
Kim et al. [126]	43.34	0.9694	47.75	0.9851	47.12	0.9838
Yang et al. [128]	43.35	0.9704	48.24	0.9868	47.45	0.9855
Sparse interpolation [129]	42.01	0.9598	44.28	0.9713	44.69	0.9739
ASDS [80]	41.69	0.9549	45.43	0.9750	44.32	0.9703
CSR [84]	42.31	0.9597	46.11	0.9781	45.27	0.9752
NCSR [98]	43.75	0.9720	48.13	0.9865	47.64	0.9861

of objective quality metrics (PSNR and SSIM) are shown in Table 6.2. The performance of regularization-based SR with Sparse Maxing Estimators (SME) [117] was also examined and that showed better PSNR and SSIM indices.

We examined EB method proposed by Kim and Kwon [126] on the cropped fundus images. We chose this method since it has outperformed many state-of-the art algorithms and also because it removes blurring and ringing effects near the edges [126]. The input LR images were created by downsampling the original image by a factor of 2. Noise was not added to the downsampled image. The training set was created by randomly selecting HR generic images. The LR training images were obtained by blurring and subsampling HR images. Thus, the training set constituted a set of LR and HR image pairs. The algorithm was performed on image patches. In this method, the input LR patch was first interpolated by a factor of 2 using cubic interpolation. Next, kernel ridge regression was adopted to learn a map from input LR patch to training set HR image and LR patch pairs. The regression provided a set of candidate images. The super-resolved image was obtained by combing through candidate images based on the estimated confidences. The artifacts around the edges of the reconstructed image were removed by utilizing image

prior regularization term. Better PSNR and SSIM values are noticed in this method.

Similarly, sparse representation-based SR techniques were examined on the LR fundus image. We extracted 5×5 patches with overlap of 1 pixel between adjacent patches from the input image. The HR dictionaries and sparse coefficients were learned from both the training set HR images and LR test image. We used the method and software proposed by Yang, Wright, Huang and Ma [128] to run the simulation. In addition, ASDS [80], sparse interpolation [129], CSR [84], and the most recent NCSR [98] methods proposed by Dong et al. were also implemented on LR patches. The latter two methods introduced the centralized sparsity constraint by exploiting non-local statics. Both the local sparsity and non-local sparsity constraints are combined in this approach. The centralized sparse representation approach approximates the sparse coefficients of the LR image as closely as the original HR image does which results in better image reconstruction and hence better PSNR and SSIM indices. Figures. 6.11, 6.12 and 6.13 show the results of various single image SR approaches applied to the LR images of optic disc, macula and retinal blood vessels.

Finally, multi-frame SR techniques were examined on multiple real retinal images. A fundus camera was fixed manually on the right eye of the author and six consecutive images were taken. The participants head was fixed using a chin rest. To eliminate pupil constriction, one drop of pupil dilating agent (Tropicamide 1%) was used. The dilating agent also minimized the accommodation effects by paralyzing the ciliary muscles. Six images were acquired with small shifts and rotations due to the eye motions. The images were then cropped to obtain the desired sections. We cropped a small retinal blood vessel section of the fundus image in this study. The cropped sections were 128×128 pixels. One such section is shown in Figure 6.14 (top left corner).

In the Fourier domain, the shifts between the images were estimated from the central 5% of the frequency and the rotations were estimated from a disc of radius 0.6 as described by Vandewalle et al. [111]. The images were registered in the Fourier domain and then reconstructed to obtain an HR image using cubic interpolation. The resolution was increased by a factor of two. The robust regularization, IBP and POCS were also tested on the real images. The results showed that cubic interpolation and POCS methods estimated the blur HR images. These images were worse than the image estimated by a single image bicubic interpolation method while the HR images estimated by IBP and robust regularization had better visual quality.

A hierarchical Bayesian algorithm was used to estimate the registration parameters in Bayesian methods and the blur was assumed to be a 3×3 Gaussian with variance 1. We examined TV-prior, l_1-prior and SAR-prior models; however, these models led to over-

smooth non-edge regions of the image. Therefore, combined TV-SAR and l_1-SAR prior models were used as described by Villena et al. [121]. The HR images obtained from these methods are shown in Figure 6.14. The image obtained from single image bicubic interpolation is also shown for comparison. The HR images suffered from registration errors more likely due to the fact that most of the SR algorithms work only with planar motion and rotations. The real images especially retinal images taken from a fundus camera may have non-planar motions due to the eye motion. A more precise knowledge of motion parameters is needed to solve this problem. The registration error may also be minimized by taking images from dynamic instruments. Last but not least, we examined blind deconvolution-based SR approach to combine multiple LR images to estimate an HR image as described by Sroubek and Flusser [122]. This method does not assume any prior information about the blur; it requires only the approximate size of the blur. In our case we set 3×3 blur kernel size. The algorithm built and minimized a regularized energy function given in equation (6.6) with respect to original image and blur. The regularization is conducted in both the image and blur domain. Total variation (TV) regularization was used for our simulation. The HR image predicted by a blind deconvolution method showed a more even spread of the brightness and the edges are sharper and clearer. Among all of the above multi-frame SR techniques, blind deconvolution showed the best results for our LR images.

6.5 Discussion and Conclusion

In this chapter, we demonstrated the possibility of resolution enhancement of retinal images using SR techniques. Three important sections of the fundus images, namely optic disc, macula and retinal blood vessels were considered for the simulations. In the first part, we simulated 4 LR images by shifting, rotating, downsampling and adding Gaussian noise to an HR image. A number of important features of the fundus image are missing or less clearly visible in the LR images. For example, the arteries and veins emerging from the margins of the optic disc are less clearly visible. The connections between the optic disc and some of the blood vessels are not distinct in the LR optic disc images. The borderlines of blood vessels are less distinct in retinal blood vessels image. Similarly, there are several tiny arteries (cilioretinal arteries) that supply additional blood to the macular region that are also less clearly visible in LR macula images. In addition, the background of the retina is unclear and therefore it is difficult to identify the cilioretinal arteries. The foveal reflex is considerably dim in LR macula images. We applied SR techniques to the images of a healthy normal eye in this thesis; the application of SR techniques might

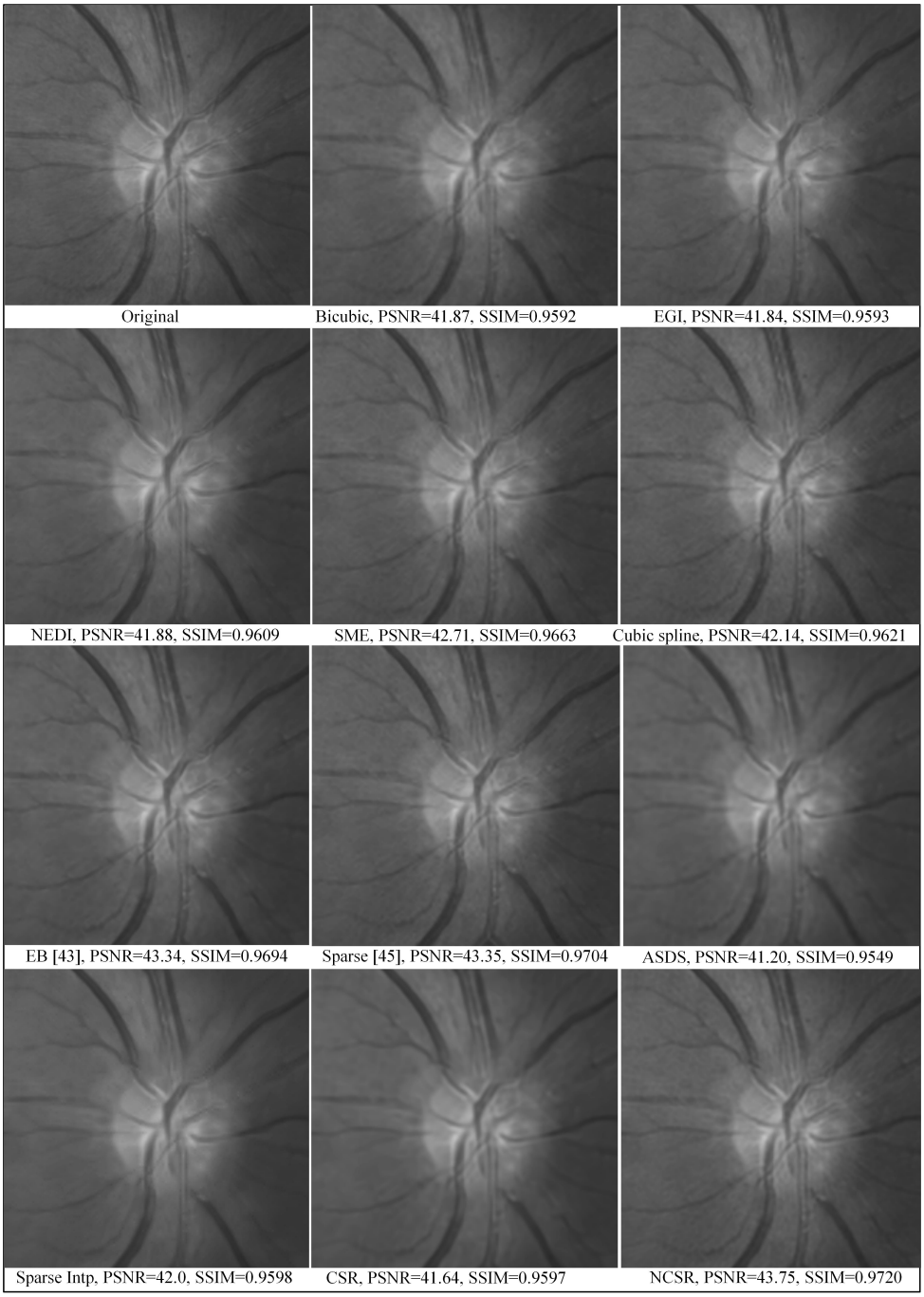


Figure 6.11: Results from various single-image SR approaches for LR optic disc images.

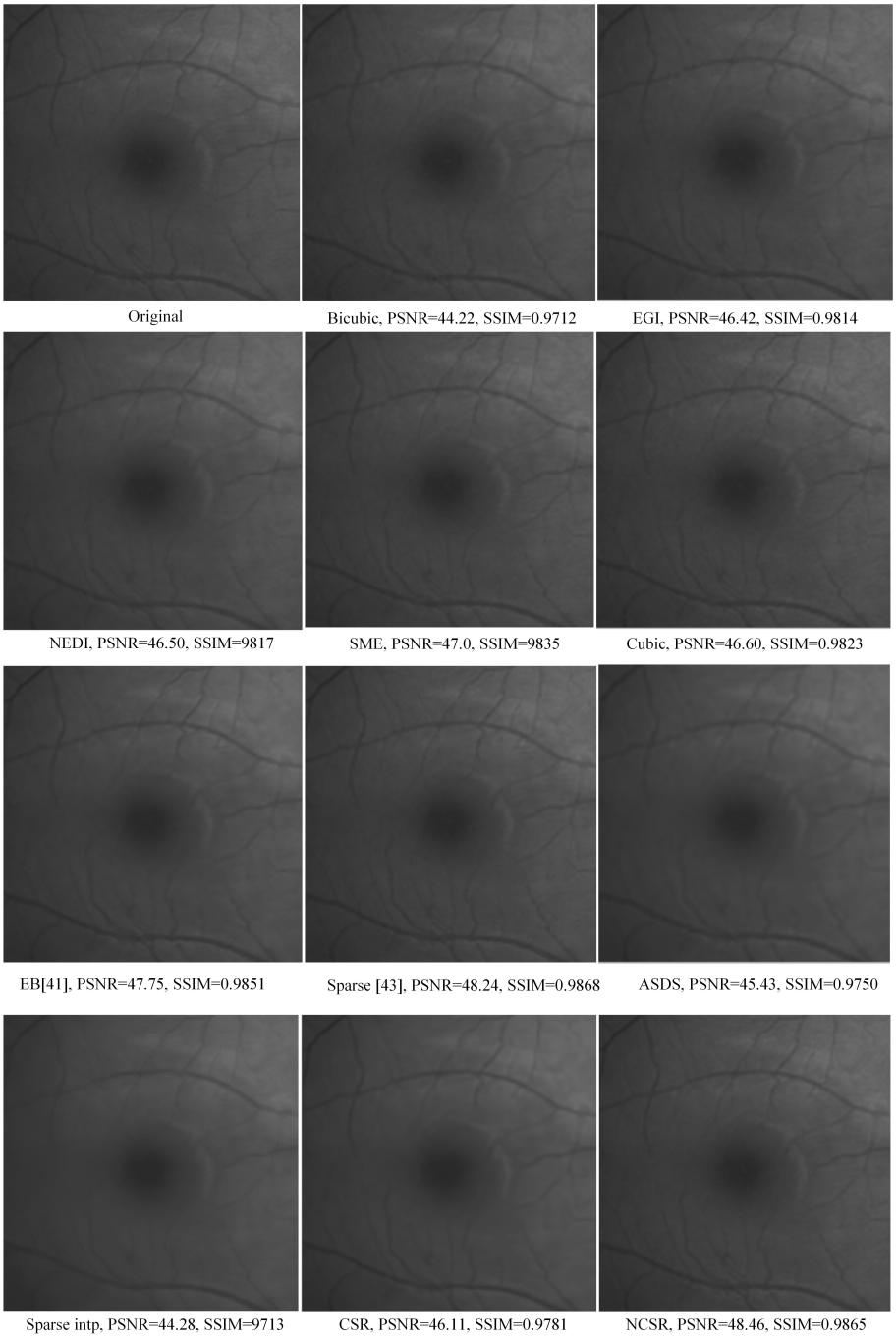


Figure 6.12: Results from various single-image SR approaches for LR macula images.

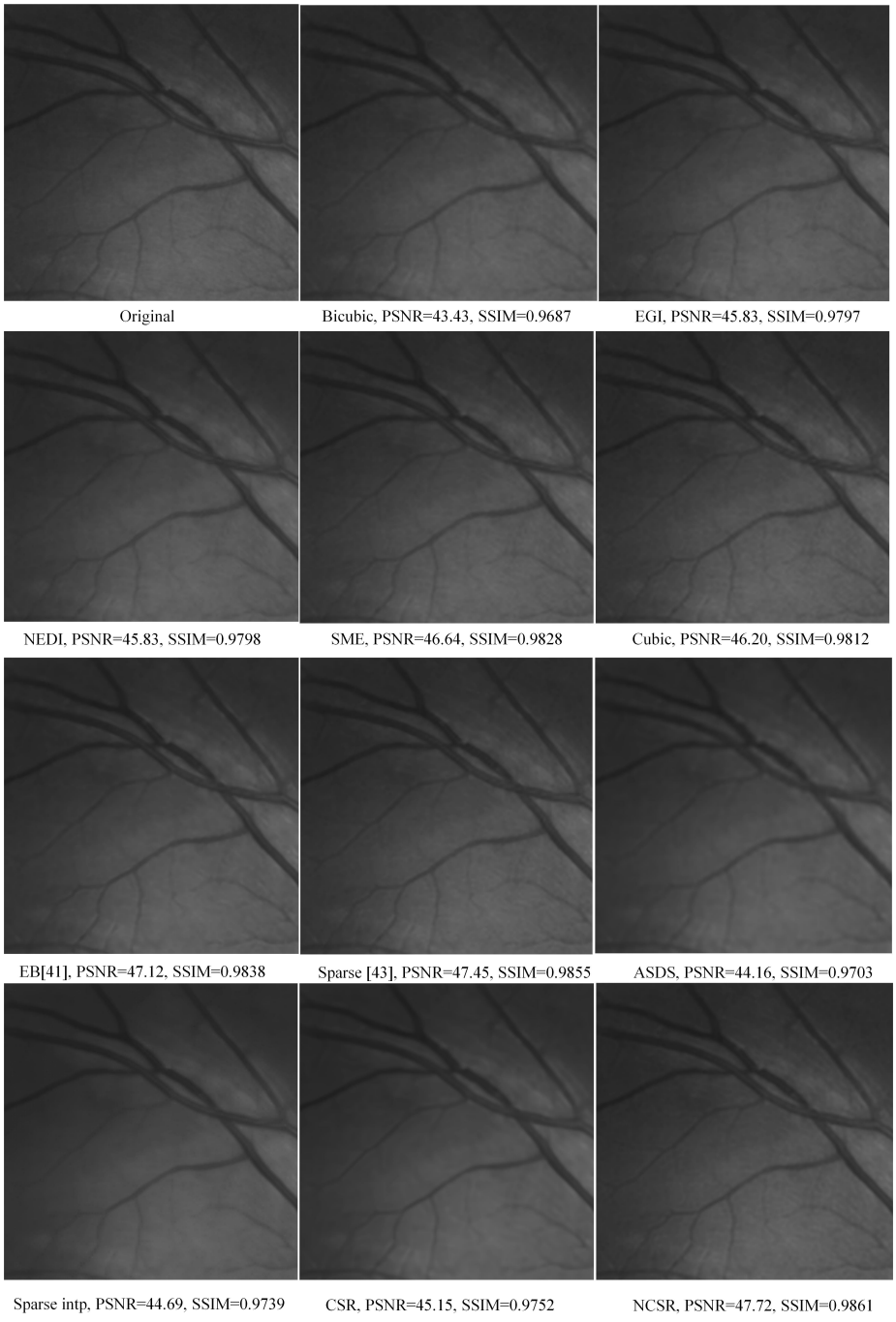


Figure 6.13: Results from various single image SR approaches for LR blood vessel images.

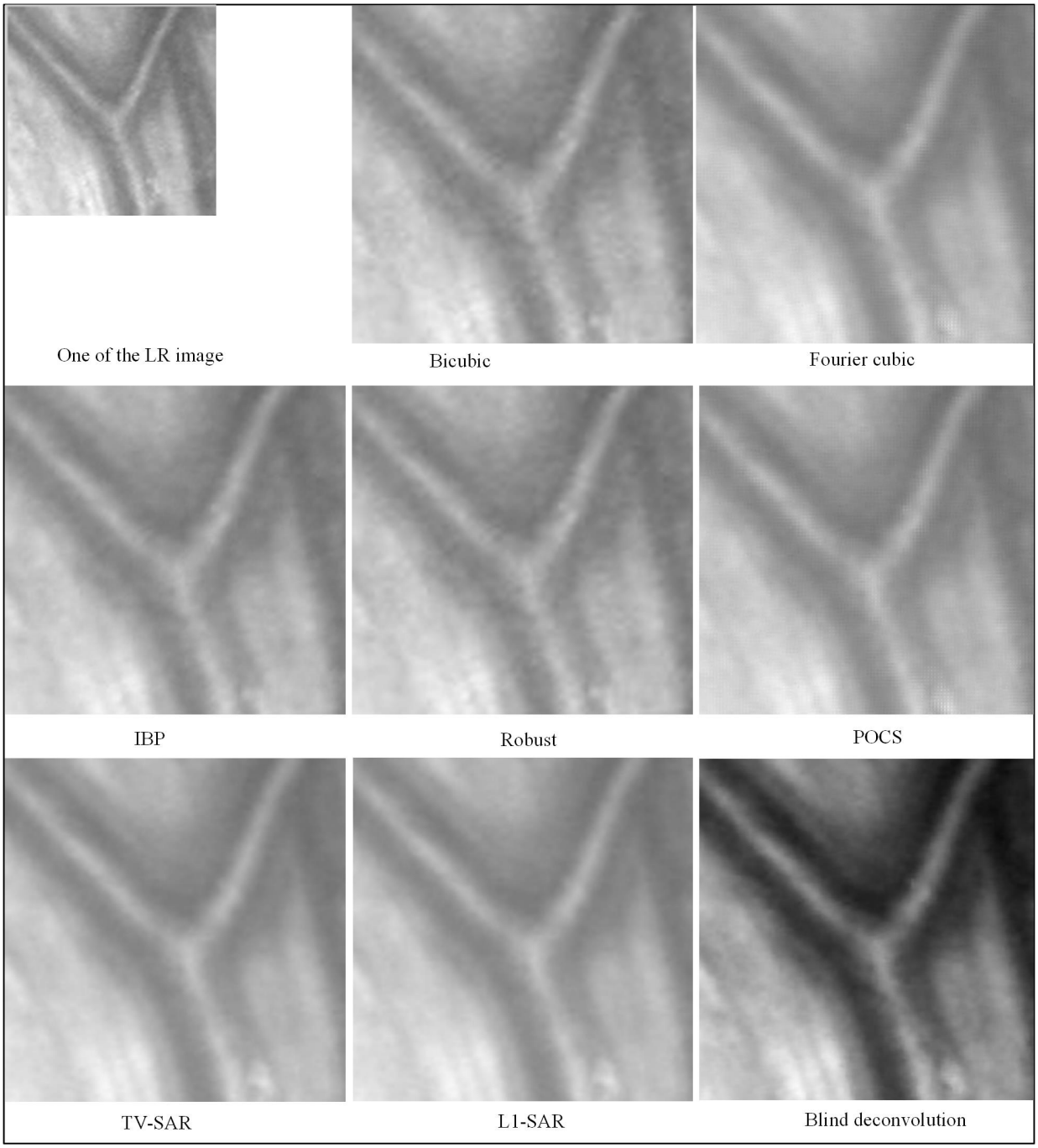


Figure 6.14: Results from various multiframe SR approaches for real retinal images.

be more important for the images taken from pathological eyes, such as DR and AMD. Microaneurysms, hemorrhages, exudates and cotton wool spots, drusen, and abnormal and fragile new blood vessels are all indicators of retinal diseases. The retinal lesions associated with retinal diseases can be enhanced by using SR techniques which may help to improve early diagnosis of retinal diseases.

SR techniques give magnified image by fusing multiple LR images. The features of an image are also magnified in the same factor so they are visually clearer in a larger image. This can be seen in many super-resolved images shown in figures 6.8 to 6.13. The features which are visually less distinct in LR images are more distinct in super-resolved image and the connection between the optic disc and blood vessels are comparatively clearer in super-resolved images given by SME [117], EB [126] and sparse representation [128] methods (Figure 6.11). However, some SR techniques such as sparse interpolation and ASDS over-smooth the image, and therefore the important features are lost from the image. The cilioretinal arteries and foveal reflex in super-resolved images given by EB [126], SME [117], and sparse [128] methods are as clear as in the original image while those in super-resolved images obtained from sparse interpolation and ASDS [80] methods are significantly smoothed out (Figure 6.12). The other SR techniques show intermediate performance. While many features have been recovered in the super-resolved images, the edges or textures lost during the decimation process could not be recovered completely. The EB [126] and sparse representation [128] methods show good results and can be used when sufficient number of input LR images are unavailable and/or when a higher resolution factor is required. The performance of these algorithms may be improved using a larger set of training images and more relevant learning method.

In the second part, we implemented SR techniques on cropped version of 6 multiple acquisition retinal images taken with a fundus camera. The real images were first normalized to reduce the effect of different levels of illumination in different images. The Fourier method [111] was used to estimate the registration parameters and these parameters were subsequently used to fuse LR images. The Fourier-based cubic interpolation method significantly blurred the reconstructed image. The IBP, robust regularization and single image bicubic interpolation method introduced small amount of ringing effect; however, they preserved most of the image features. The Bayesian approaches performed registration and fusion tasks simultaneously. They provided visually pleasant images; however, the HR images reconstructed by the Bayesian approaches were over-smoothed and many image details were lost. The blind deconvolution approach provided much sharper and cleaner reconstructed image than others; however, the HR image was not free of artifacts. Nevertheless, the algorithm is more realistic because it doesn't need prior information of the blur. In the real image SR approaches we utilized only 6 LR images to perform SR; better results

can be obtained if the number of LR image is increased. Meitav and Ribak [130] used more than 200 retinal LR images to achieve an HR image using a weighted average method. Furthermore, the registration methods are restricted to the planar translational and rotational while the eye movements may induce non-planar variations between LR images. The variations in the ocular surface such as the dynamic nature of the tear film, accommodation and cardiopulmonary effect of the eye may yield variations in the LR retinal images. Small sub-pixel error in the registration may result in different estimation. Therefore, a robust motion estimation algorithm is essential to perform SR in retinal images. Furthermore, many algorithms including Bayesian approaches assume spatially uniform Gaussian blur which is usually impractical. To avoid damage to the eye, the retinal images are taken in low light condition, and therefore suffer from low signal to noise ratio. Most of the SR algorithms deteriorate when noise is present in the image; therefore, a method which is more robust to noise while preserving image features is essential.

6.6 Summary

In summary, we demonstrated resolution enhancement of retinal images using image processing methods. While we investigated the performance of various SR techniques, we are unable to present the details of each method in this thesis. The reader is advised to refer to the related reference papers for specific implementation details. Since the codes were downloaded from the websites of the respective authors and the parameters of each method were set according to the values given in their corresponding papers, the differences in PSNRs and SSIMs between various SR approaches may be due to the differences in techniques, and/or their parameters. We refer the interested reader to visit the webpage http://quark.uwaterloo.ca/ dthapa which contains the MATLAB source code for various SR techniques developed by several groups of researchers.

Chapter 7

Conclusion and Future Directions

7.1 Conclusions

This dissertation discusses the application of image restoration methods on retinal images. The dissertation begins with introducing imaging instruments such as OCT, and the quality of the images provided by imaging instruments. A raw image taken with an imaging instrument usually goes through a post-processing stage. The purpose of post-processing is to improve the image quality. Image denoising and interpolation are two important post-processing techniques applied to retinal images. In this thesis, we developed new denoising and interpolation algorithms for OCT images that outperform several existing algorithms.

Sparse representation is the most compact representation of a signal. The effectiveness of the signal representation depends mostly on the choice of dictionary. A dictionary can be created in two ways. The analytical-based dictionaries are created by changing the parameters of a known mathematical function, while the learning-based dictionaries are trained from a training set. Analytical-based dictionaries are simple and can be created very fast; however, their performance depends on the nature of the signal. There is not a single dictionary that works well for all kinds of signals. On the other hand, trained dictionaries are made for a specific signal, and therefore perform better than the analytical-based dictionary. Nonetheless the trained dictionary is not flexible and the computational complexity of training a dictionary is inevitable. We created a new analytical-based dictionary (nonlinear dictionary) that addresses these challenges to some extent. The proposed dictionary performed better than DCT and Haar dictionaries and does not require a time consuming learning process. The proposed dictionary is a hybrid dictionary which consists of a diverse set of atoms; therefore, it provides less reconstruction error not only for

harmonic and non-harmonic signal reconstructions but also for image reconstruction. We implemented our NL dictionary for fundus and OCT image reconstruction. The results showed that image reconstruction improved (in terms of PSNR) with the addition of each set of NL atoms. The NL dictionary performed better than DCT and Haar dictionaries for fundus and OCT image reconstruction; however, its performance was slightly poorer than the trained dictionary. The NL dictionary is appropriate for medical images where the quality of the reconstructed image is very important. KSVD dictionary provides the best quality images; however, it needs high storage space and therefore it is not suitable for medical images. Hospitals need to store lots of images from the patients, sometimes for a very long time to allow proper follow up. The NL dictionary on the other hand is suitable for applications where lots of images need to be stored. Similarly NL dictionary can be used for telemedicine. Speckle noise is inherent in OCT images. In chapter 4, we utilized our NL dictionary for speckle noise reduction in SD-OCT images. The denoising method uses B-scans taken from a nearby location of SD-OCT volume, therefore, it does not require multiple scanning of identical locations. The aforementioned significantly reduces the scanning time. B-scans from nearby locations of SD-OCT volumes are registered and simultaneously denoised using a sparsity-based approach. Finally the median pixels of the denoised patches from the nearby frames are chosen to reconstruct an image that further enhances the denoising. The proposed multi-frame denoising approach is capable of reducing the speckle noise from the SD-OCT images while preserving the image features. Since there is a clear trade-off between the image feature preservation and speckle noise reduction, our denoising method provides an image without degrading the retinal structures. Although the K-SVD dictionary provides better quality images, it has a much longer computation time. Due to its inevitably long computation time, the KSVD dictionary cannot be applied in real time.

The computation time can also be reduced by utilizing a low-rank approximation. There are several applications of low-rank approximation. In this thesis however, we only explained its application in reconstruction of a noise free image from a noisy image. In our denoising method, the B-scan was divided into several small matrices consisting of non-local similar patches. Block matching algorithm was used to search for the non-local similar patches and stacked them into a matrix. Each column of the matrix consisted of noisy patches from the image. A noise free patch matrix was obtained from noisy patch matrix using WNNM. Finally the denoised patches from the nearby frames were averaged to enhance the denoising. The denoised image obtained from the proposed approach was better than those obtained by several state-of-the-art methods. In the second part of the Chapter 5, we extended the proposed denoising approach for joint denoising and interpolation of SD-OCT images. The multiple denoised frames were interpolated and fused to

get an HR image. The information of the HR image were incorporated to denoise and interpolate individual frames. We used SVD fusion method to add information from an HR image to the test image. The proposed denoising approach has several advantages. First, it does not require multiple imaging of one location. Regular and fast scanning machines can be used to acquire noisy images from an eye, significantly reducing the scanning time and hence the image artifacts due to eye motion. The LR images taken by low cost OCT can be simultaneously denoised and interpolated to acquire an HR, high SNR image.

Finally we examined the possibility of resolution enhancement of fundus images using several SR techniques. Previously written and publicly available MATLAB codes were downloaded and their performance on fundus images were examined. Three important sections of the fundus images, namely optic disc, macula and retinal blood vessels, were cropped from the fundus images and were used to study the effectiveness of the SR techniques. The SR techniques were compared with both the simulated LR images and real experimental images. A number of important features of the fundus image were missing or less clearly visible in the LR images. These features were magnified by the SR techniques so that they were clearer in the super-resolved images. However, when used in real experimental images, the SR techniques provided image artifacts. This could be due to eye movements which may induce non-planar variations between LR images. Therefore, a robust motion estimation algorithm or an eye tracking system is required to perform SR on real experimental images.

7.2 Future Directions

The NL dictionary has lots of room for improvement. It is possible that other NL atoms can further reduce signal reconstruction error; therefore, a possible future work is to look for dictionary atoms that provide better performance. We added a set of evenly selected phases between 0 to 2π to the conventional DCT atoms to create a phase added dictionary. The performance of the phase added dictionary can be improved by adding non-linear phases so that the dictionary correlates better with the signal. Besides, the NL dictionary consists of many parameters and some of them are not well defined. These parameters should be clearly defined to make generating the atoms easier and the dictionary more flexible. For example, polynomial atoms were created using predefined set of coefficients which were not optimized. The coefficient optimization may improve the performance of the NL dictionary for signal reconstruction. We can also adopt dictionary optimization techniques to optimize the atoms of the dictionary so that the size of the dictionary is reduced without degrading its performance. We applied NL dictionary for fundus and

OCT image reconstruction and image denoising. The effectiveness of the dictionary can be examined for other image restoration processes such as feature extraction, interpolation, image compression and automated detection.

The sparsity-based denoising method proposed in Chapter 4 can be further extended to image interpolation. There is a high degree of similarity between the LR and HR images. The learning technique can be used to find the matching parameters between the test LR image and the training set HR images. The matching parameters can be incorporated to the LR image to increase their resolution. The efficiency of denoising can be increased by training a KSVD dictionary from HR images from the same location of retina.

In Chapter 5, we proposed a joint denoising and interpolation technique for SD-OCT images. In this technique we implemented simple bicubic interpolation method to fuse multiple images since this method is very easy and computationally fast. A better SR technique can be employed to improve the results.

Chapter 6 describes a variety of techniques for SR of fundus images. We implemented super-resolution techniques on gray-scale images. The micro scale retinal structures less clearly visible or missing in the LR images were enhanced in the super-resolved images. These structures were magnified in the super-resolved images to enhance visibility. However, fundus color also help clinicians to detect early stages of retinal diseases. Therefore, another future work would be applying SR techniques on fundus color images. Moreover, the images used in chapter 6 were taken from healthy individuals. It would be interesting to examine the performance of the SR techniques on images of patients to verify the clinical or diagnostic value of these software. Findings of such studies could improve diagnosis of retinal diseases.

Bibliography

[1] H. Jelinek and M. Cree, "Introduction," in *Automated image detection of retinal pathology* (H. Jelinek and M. Cree, eds.), pp. 1–13, CRC Press, Boca Raton, 2010.

[2] K. Yogesan, F. Reinholz, and I. J. Constable, "Tele-diabetic retinopathy screening and imaged -based clinical decision support," in *Automated image detection of retinal pathology* (H. Jelinek and M. Cree, eds.), pp. 339–348, CRC Press, Boca Raton, 2010.

[3] J. Fujimoto, C. Pitris, S. Boppart, and M. Brezinski, "Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy," *Neoplasia*, vol. 2, no. 1, pp. 9–25, 2000.

[4] J. S. Schuman, "Spectral domain optical coherence tomography for glaucoma (an aos thesis)," *Transactions of the American Ophthalmological Society*, vol. 106, pp. 426–458, 2008.

[5] S. A. Boppart, "Optical coherence tomography: technology and applications for neuroimaging," *Psychophysiology*, vol. 40, no. 4, pp. 529–541, 2003.

[6] J. Izatt and M. Choma, "Theory of optical coherence tomography," in *Optical Coherence Tomography*, pp. 47–72, Springer, Berlin, Heidelberg, 2008.

[7] X. Zha, "Image processing of optical coherence tomography image guided dental drilling," Master's thesis, Department of Biomedical engineering, Delft University of Technology, Netherlands, 2011.

[8] F. Gasca, L. Ramrath, G. Huettmann, and A. Schweikard, "Automated segmentation of tissue structures in optical coherence tomography data," *Journal of biomedical optics*, vol. 14, no. 3, pp. 034046–034046, 2009.

[9] J. Schmitt, S. Xiang, and K. Yung, "Speckle in optical coherence tomography: an overview," in *Proc. SPIE 3726, Saratov Fall Meeting '98: Light Scattering Technologies for Mechanics, Biomedicine, and Material Science*, pp. 450–461, 1999.

[10] L. Fang, S. Li, R. McNabb, Q. Nie, A. Kuo, C. A. Toth, J. Izatt, and S. Farsiu, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE transactions on medical imaging*, vol. 32, no. 11, pp. 2034–2049, 2013.

[11] D. Thapa, K. Raahemifar, and V. Lakshminarayanan, "A new efficient dictionary and its implementation on retinal images," in *Proc. IEEE 19th International Conference on Digital Signal Processing*, pp. 841–846, 2014.

[12] D. Thapa, K. Raahemifar, and V. Lakshminarayanan, "Performance analysis of unconventional dictionary on retinal images," in *Proc. IEEE International Conference image processing*, pp. 2275–2279, 2014.

[13] D. Thapa, K. Raahemifar, and V. Lakshminarayanan, "Reduction of speckle noise from optical coherence tomography images using multi-frame weighted nuclear norm minimization method," *In Press*.

[14] D. Thapa, K. Raahemifar, W. Bobier, and V. Lakshminarayanan, "Comparison of super-resolution algorithms applied to retinal images," *Journal of biomedical optics*, vol. 19, no. 5, pp. 056002–056002, 2014.

[15] G. A. Campbell and R. M. Foster, *Fourier integrals for practical applications*. New York: Bell telephone laboratories, 1948.

[16] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. 100, pp. 90–93, 1974.

[17] I. Daubechies, *Ten lectures on wavelets*. Philadelphia: Society for industrial and applied mathematics, 1992.

[18] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, pp. 713–718, 1992.

[19] D. L. Donoho, "Wedgelets: Nearly minimax estimation of edges," *The Annals of Statistics*, vol. 27, pp. 859–897, 1999.

[20] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in *Proc. IEEE Digital Signal Processing Workshop*, pp. 120–131, 1998.

[21] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling and Simulation*, vol. 5, pp. 861–899, 2006.

[22] E. Le-Pennec and S. Mallat, "Bandelet image approximation and compression," *Multiscale Modeling and Simulation*, vol. 4, pp. 992–1039, 2005.

[23] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, pp. 2091–2106, 2005.

[24] Y. Lu and M. N. Do, "3-d directional filter banks and surfacelets," in *Proc. SPIE Optics and Photonics* , pp. 59141Q–59141Q, 2005.

[25] I. Jolliffe, *Principal component analysis*. John Wiley and Sons, 2002.

[26] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, pp. 287–314, 1994.

[27] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and computational harmonic analysis*, vol. 21, pp. 5–30, 2006.

[28] K. Engan, S. O. Aase, and J. Husoy, "Frame based signal compression using method of optimal directions (mod)," in *Proc. IEEE International Symposium on Circuits and Systems* , pp. 1–4, 1999.

[29] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4311–4322, 2006.

[30] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Applied and computational harmonic analysis*, vol. 21, pp. 53–94, 2006.

[31] R. Kafieh, H. Rabbani, and I. Selesnick, "Three dimensional data-driven multi scale atomic representation of optical coherence tomography," *IEEE Transactions on Medical Imaging, In Press*, 2014.

[32] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[33] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[34] G. Rath and A. Sahoo, "A comparative study of some greedy pursuit algorithms for sparse approximation," in *Proc. Third Annual Symposium on Combinatorial Search*, vol. 19, pp. 37–53, 2009.

[35] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[36] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *Technical Report-CS Technion*, 2008.

[37] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[38] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[39] K. Kreutz-Delgado, J. Murray, B. Rao, K. Engan, T. Lee, and T. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computation*, vol. 15, no. 2, pp. 349–396, 2003.

[40] M. Marsousi, "Variable length k-svd: A new dictionary learning approach and multistage omp method for sparse representation," Master's thesis, Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada, 2012.

[41] R. D. Dony, "Karhunen-loeve transform," in *The Transform and Data Compression Handbook* (K. R. Rao and P. Yip, eds.), CRC Press, Boca Raton, 2001.

[42] A. d'Aspremont, L. E. Ghaoui, M. I. Jordan, and G. R. G. Lanckriet, "A direct formulation for sparse pca using semidefinite programming," *SIAM Review*, vol. 49, no. 3, pp. 434–448, 2007.

[43] Y. Peng, J. L. Suo, Q. Dai, and W. Xu, "From compressed sensing to low-rank matrix recovery: Theory and applications," *Acta Automatica Sinica*, vol. 39, no. 7, pp. 981–994, 2013.

[44] E. Cands, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[45] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[46] E. J. Candes and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decompositions," *Foundations of Computational Mathematics*, vol. 6, no. 2, pp. 227–254, 2006.

[47] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[48] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[49] D. Thapa, K. Raahemifar, and V. Lakshminarayanan, "Less is more: compressive sensing in optics and image science," *Journal of Modern Optics*, pp. 1–15, 2014.

[50] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[51] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2862–2869, 2014.

[52] J. F. Cai, E. Cands, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[53] D. Zhang, Y. Hu, J. Ye, X. Li, and X. He, "Matrix completion by truncated nuclear norm regularization," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2192–2199, 2012.

[54] Z. Wang, A. Bovik, H. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[55] S. G. M. G. M. Mrak, "Reliability of objective picture quality measures," *Journal of Electrical Engineering*, vol. 55, no. 1-2, pp. 3–10, 2004.

[56] M. Tharmalingam, "Dictionaries and algorithms for sparsity constrained image reconstruction," Master's thesis, Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada, 2013.

[57] R. Rubinstein, "http://www.cs.technion.ac.il/ ronrubin/software.html," Accessed on December 2014.

[58] L. Bian, J. Suo, F. Chen, and Q. Dai, "Denoising high speed optical coherence tomography data using inter-frame and intra-frame priors," *arXiv:1312.1931*, 2013.

[59] M. Szkulmowski, I. Gorczynska, D. Szlag, M. Sylwestrzak, A. Kowalczyk, and M. Wojtkowski, "Efficient reduction of speckle noise in optical coherence tomography," *Optics express*, vol. 20, no. 2, pp. 1337–1359, 2012.

[60] Y. Huang and J. van Genderen, "Evaluation of several speckle filtering techniques for ers-1&2 imagery," *International Archives of Photogrammetry and Remote Sensing*, vol. 31, pp. 164–169, 1980.

[61] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 165–168, 1996.

[62] D. T. Kuan, A. Sawchuk, T. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 165–177, 1985.

[63] A. Aboshosha, M. Hassan, and M. E. Mashade, "Image denoising based on spatial filters, an analytical study," in *Proc. IEEE International Conference on Computer Engineering and Systems*, pp. 245–250, 2009.

[64] V. Frost, J. Stiles, K. Shanmugan, and J. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 157–166, 1982.

[65] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[66] M. J. Black, G. Sapiro, D. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 421–432, 1998.

[67] A. Buades, B. Coll, and J. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 60–65, 2005.

[68] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, pp. 613–627, 1995.

[69] S. Sudha, G. Suresh, and R. Suknesh, "Speckle noise reduction in ultrasound images by wavelet thresholding based on weighted variance," *International Journal of Computer Theory and Engineering*, vol. 1, no. 1, pp. 7–12, 2009.

[70] S. Sudha, G. Suresh, and R. Suknesh, "Speckle noise reduction in ultrasound images using context-based adaptive wavelet thresholding," *IETE Journal of Research*, vol. 55, no. 3, pp. 135–143, 2009.

[71] D. L. Donoho and J. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[72] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, pp. 1532–1546, 2000.

[73] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, pp. 1338–1351, 2003.

[74] A. Wong, A. Mishra, K. Bizheva, and D. Clausi, "General bayesian estimation for speckle noise reduction in optical coherence tomography retinal imagery," *Optics Express*, vol. 18, pp. 8338–8352, 2010.

[75] M. Welling and P. Gehler, "Products of edge-perts," in *Proc. Advances in neural information processing systems*.

[76] Z. Jian, L. Yu, B. Rao, B. Tromberg, and Z. Chen, "Three-dimensional speckle suppression in optical coherence tomography based on the curvelet transform," *Optics Express*, vol. 18, no. 2, pp. 1024–1032, 2010.

[77] P. Hiremath, P. Akkasaligar, and S. Badiger, "Speckle reducing contourlet transform for medical ultrasound images," *Int J Compt Inf Engg*, vol. 4, no. 4, pp. 284–291, 2010.

[78] W. Dong, G. Shi, and X. Li, "Image deblurring with low-rank approximation structured sparse representation," in *Proc. Signal and Information Processing Association Annual Summit and Conference*.

[79] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, pp. 1531–1549, 2010.

[80] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838–1857, 2011.

[81] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, pp. 3736–3745, 2006.

[82] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, pp. 2080–2095, 2007.

[83] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*.

[84] W. Dong, D. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proc. IEEE International Conference on image restoration*.

[85] A. Ozcan, A. Bilenca, A. Desjardins, B. Bouma, and G. Tearney, "Speckle reduction in optical coherence tomography images using digital filtering," *J. Opt. Soc. Am. A*, vol. 24, no. 7, pp. 1901–1910, 2007.

[86] B. Sander, M. Larsen, L. Thrane, J. Hougaard, and T. Jrgensen, "Enhanced optical coherence tomography imaging by multiple scan averaging," *British Journal of Ophthalmology*, vol. 89, no. 2, pp. 207–212, 2005.

[87] A. Sakamoto, M. Hangai, and N. Yoshimura, "Spectral-domain optical coherence tomography with multiple b-scan averaging for enhanced imaging of retinal diseases," *Ophthalmology*, vol. 115, no. 6, pp. 1071–1078, 2008.

[88] M. Mayer, A. Borsdorf, M. Wagner, J. Hornegger, C. Mardin, and R. Tornow, "Wavelet denoising of multiframe optical coherence tomography data," *Biomedical optics express*, vol. 3, no. 3, pp. 572–589, 2012.

[89] S. Chitchian, M. Mayer, A. Boretsky, F. van Kuijk, and M. Motamedi, "Retinal optical coherence tomography image enhancement via shrinkage denoising using double-density dual-tree complex wavelet transform," *J. Biomed. Opt*, vol. 17, no. 11, pp. 116009(1)–116009(4), 2012.

[90] L. Fang, S. Li, Q. Nie, J. Izatt, C. Toth, and S. Farsiu, "Sparsity based denoising of spectral domain optical coherence tomography images," *Biomedical optics express*, vol. 3, no. 5, pp. 927–942, 2012.

[91] A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Process*, vol. 86, no. 3, pp. 572–588, 2006.

[92] L. Fang, "http://people.duke.edu/∼sf59/fang_tmi_2013.htm," Accessed on May, 10th 2014.

[93] L. Zhang, "http://www4.comp.polyu.edu.hk/∼cslzhang/," Accessed on May, 10th 2014.

[94] S. Mallat, "Wavelets for a vision," in *Proc. IEEE*, vol. 84, pp. 604–614, 1996.

[95] M. D. Robinson, C. A. Toth, J. Y. Lo, and S. Farsiu, "Efficient fourier-wavelet super-resolution," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2669–2681, 2010.

[96] L. Zhang, X. Li, and D. Zhang, "Image denoising and zooming under the lmmse framework," *IET Image Processing*, 2010.

[97] H. Nasir, V. Stankovi, and S. Marshall, "Singular value decomposition based fusion for super-resolution image reconstruction," *Signal Processing: Image Communication*, vol. 27, no. 2, pp. 180–191, 2012.

[98] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, 2013.

[99] J. Tian and K. Ma, "A survey on super-resolution imaging," *Signal, Image and Video Processing*, vol. 5, no. 3, pp. 329–342, 2011.

[100] D. Keren, S. Peleg, and R. Brada, "Image sequence enhancement using sub-pixel displacements," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 742–746, 1988.

[101] J. R. Bergen, P. A. ans K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proc. 2nd European Conference on Computer Vision, Lecture Notes in Computer Science*, pp. 237–252, 1992.

[102] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.

[103] D. Capel and A. Zisserman, "Computer vision applied to super-resolution," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 75–86, 2003.

[104] R. Hardie, K. J. Barnard, and E. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.

[105] H. Ur and D. Gross, "Improved resolution from sub-pixel shifted pictures," *CVGIP Graph Models and Image Processing*, vol. 54, pp. 181–186, 1992.

[106] X. Li and M. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521–1527, 2001.

[107] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2226–2238, 2006.

[108] C. Liu and D. Sun, "A bayesian approach to adaptive video super resolution," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 209–221, 2011.

[109] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *JOSA A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[110] R. Y. Tsai and T. Huang, "Multiframe image restoration and registration," *Advances in Computer Vision and Image Processing*, vol. 1, no. 2, pp. 317–339, 1984.

[111] P. Vandewalle, S. S, and M. Vetterli, "A frequency domain approach to registration of aliased images with application to super-resolution," *EURASIP Journal on Applied Signal Processing*, pp. 1–14, 2006.

[112] S. Rhee and M. Kang, "Discrete cosine transform based regularized high-resolution image reconstruction algorithm," *Optical Engineering*, vol. 38, no. 8, pp. 1348–1356, 1999.

[113] N. Nguyen and P. Milanfar, "A wavelet-based interpolation-restoration method for super-resolution (wavelet super-resolution)," *Circuits, Systems, and Signal Processing*, vol. 19, no. 4, pp. 321–338, 2000.

[114] E. Plenge, D. Poot, M. Bernsen, G. Kotek, G. Houston, P. Wielopolski, L. van der Weerd, W. J. Niessen, and E. Meijering, "Super-resolution reconstruction in mri: better images faster?," in *Proc. SPIE 8314*, pp. 138.1–138.8, 2012.

[115] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[116] A. Zomet, A. Rav-Acha, and S. Peleg, "Robust super-resolution," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. I645–I650, 2001.

[117] S. Mallat and G. Yu, "Super-resolution with sparse mixing estimators," *IEEE Transactions on Image processing*, vol. 19, no. 11, pp. 2889–2900, 2010.

[118] S. Borman and R. Stevenson, "Super-resolution from image sequences-a review," in *Proc. Midwest Symposium on Circuits and Systems*, pp. 374–378, 1998.

[119] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Total variation super resolution using a variational approach," in *Proc. IEEE International Conference on Image Processing*, pp. 641–644, 2008.

[120] S. Villena, M. Vega, R. Molina, and A. K. Katsaggelos, "Bayesian super-resolution image reconstruction using an l1 prior," in *Proc. 6th International Symposium on Image and Signal Processing and Analysis*, pp. 152–157, 2009.

[121] S. Villena, M. Vega, S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian combination of sparse and non sparse priors in image super resolution," *Digital Signal Processing*, vol. 23, no. 2, pp. 530–541, 2013.

[122] F. roubek and J. Flusser, "Multichannel blind deconvolution of spatially misaligned images," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 874–883, 2005.

[123] F. roubek and J. Flusser, "Resolution enhancement via probabilistic deconvolution of multiple degraded images," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 287–293, 2006.

[124] W. Wu, Z. Liu, W. Gueaieb, and X. He, "Single-image super-resolution based on markov random field and contourlet transform," *Journal of Electronic Imaging*, vol. 20, no. 2, pp. 023005–023005, 2011.

[125] W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[126] K. Kim and Y. Kwon, "Example-based learning for single-image super-resolution and jpeg artifact removal," *Max Planck Institute Biological Cybernetics, Tbingen, Germany, Technical Report No. TR-173*, 2008.

[127] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

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

[129] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1382–1394, 2013.

[130] N. Meitav and E. N. Ribak, "Improving retinal image resolution with iterative weighted shift-and-add," *JOSA A*, vol. 28, no. 7, pp. 1395–1402, 2011.

APPENDICES

Copyright Permissions

Copyright: Permission to use images

Re: Permission to use the dataset
Sina Farsiu [farsiu@gmail.com]
Sent: Tuesday, March 04, 2014 1:26 PM
To: Damber Thapa

Hi Damber,
As far as you cite our corresponding articles, you are most welcome to use any image on my website in any form that you wish in dissertation, papers, presentations, etc.

Cheers,
-s

Sina Farsiu, Ph.D.
Assistant Professor: Departments of Ophthalmology, Biomedical Engineering, Electrical & Computer Engineering, and Computer Science
Director of the Vision and Image Processing (VIP) Laboratory Duke University
Associate Editor: IEEE Transactions on Image Processing
http://www.duke.edu/∼sf59/
Phone: (919)-684-6642

On Mar 4, 2014, at 11:04 AM, Damber Thapa <damber.thapa@uwaterloo.ca> wrote:

Dear Dr. Farsiu,
I'm a PhD candidate at the School of Optometry and Vision Science, University of Waterloo. I'm in the process of preparing my dissertation and am seeking permission to include the Optical Coherence Tomography (OCT) images that you have published in the following link:

http://people.duke.edu/∼sf59/Fang_TMI_2013.htm

The images will be used to test the OCT image restoration methods that are being developed in our lab. A credit line acknowledging the original source will be included in my dissertation. If you need any further information to help process this request, then I would appreciate it if you could let me know.

Thank you very much for your consideration.

Sincerely,
Damber Thapa

Damber Thapa (Graduate Student)
University of Waterloo
1(519)888-4567 ext. 36210
dthapa@uwaterloo.ca

Copyright: Permission to reproduce the material

RE: Permission to reproduce the material
Karen Thomas [karent@spie.org]
Sent: Thursday, April 09, 2015 12:56 PM
To: Damber Thapa

Dear Damber Thapa,
Thank you for seeking permission from SPIE to reprint material from our publications. As author, SPIE shares the copyright with you, so you retain the right to reproduce your paper in part or in whole.
Publisher's permission is hereby granted under the following conditions:
(1) the material to be used has appeared in our publication without credit or acknowledgment to another source; and
(2) you credit the original SPIE publication. Include the authors' names, title of paper, volume title, SPIE volume number, and year of publication in your credit statement.

Sincerely,
Karen Thomas for
Eric Pepper, Director of Publications
SPIE
P.O. Box 10, Bellingham WA 98227-0010 USA
360/676-3290 (Pacific Time) eric@spie.org

From: Damber Thapa [mailto:damber.thapa@uwaterloo.ca]
Sent: Tuesday, April 7, 2015 8:02 AM
To: reprint_permission
Subject: Permission to reproduce the material

Dear Editorial Office,
I hereby request permission to reproduce the following paper which has previously been published in the Journal of Biomedical Optics for inclusion in my Doctoral Thesis at the University of Waterloo.

Paper details: Damber Thapa, Kaamran Raahemifar, William R. Bobier, Vasudevan Lakshiminarayanan, "Comparison of super-resolution algorithms applied to retinal images." J. Biomed. Opt. 19 (5), 056002 (2014).
doi: 10.1117/1.JBO.19.5.056002

The original article will be cited within the thesis. I should also note that the thesis may be made electronically available to the public.

I thank you for your consideration, and look forward to hearing form you.

Sincerely,
Damber Thapa

Damber Thapa
Ph.D. Candidate
School of Optometry and Vision Science, University of Waterloo
200 University Avenue West, Waterloo, ON, N2L 3G1, Canada
Tel: +1 519 888 4567 ext. 36210