An Optimization-based Approach to Image Fusion using Structural Similarity

by

Fatemeh Hirbodvash

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Master of Science (MSc) in Computer Science

Faculty of Science
University of Ontario Institute of Technology (Ontario Tech University)

Oshawa, Ontario, Canada

May 2023

THESIS EXAMINATION INFORMATION

Submitted by: **Fatemeh Hirbodvash**

Master of Science in Computer Science

Thesis Title: An Optimization-based Approach to Image Fusion using Structural Similarity

An oral defense of this thesis took place in May 2023 in front of the following examining committee:

Examining Committee:

Chair of Examining Committee	Dr. Pooria Madani
Research Supervisor	Dr. Mehran Ebrahimi
Examining Committee Member	Dr. Heidar Davoudi
Thesis Examiner	Dr. Samaneh Mazaheri

The above committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

Abstract

This thesis examines a framework for solving optimization problems involving structural similarity index measure Mean.

Image fusion involves combining multiple images to preserve their most desirable characteristics. The objective is to create a single enhanced image that can be utilized for various applications, including human visual perception, object detection, and target recognition. A general framework is introduced for the formulation of image fusion, which incorporates the SSIM, one of the most effective measures of visual proximity that is consistent with the human visual system.

The concept of image fusion is revisited, its importance is emphasized, and an overview of methods commonly used for image fusion is provided. An alternative expression of SSIM using Mean values and vector norms is proposed, along with the calculation of its derivative. Next, the SSIM Mean is used to formulate image fusion as data fidelity of an optimization problem. Gradient-based methods are applied to solve it.

A comparison is conducted between the proposed method and the method presented by Brunet in his Ph.D. thesis at UWaterloo. The experiment use test images in MATLAB. The evaluation metrics MSE, PSNR, and SSIM show expected results in assessing the experimental outcomes. The significant advantage of the proposed formulation is its flexibility in accommodating different degradation operators and regularization terms.

Keywords: Structural Similarity Index Measure, Fusion Image, Mathematical Imaging, Numerical Optimization.

Author's Declaration

I hereby declare that this submission is entirely my own work, in my own words, and that all sources used in researching it are fully acknowledged. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize the University of Ontario Institute of Technology (Ontario Tech University) to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

_____ Fatemeh Hirbodvash

Statement of Contributions

The methodologies and results outlined in Chapters 4 and 5 are being prepared for publication.

Acknowledgements

I would like to thank Prof. Mehran Ebrahimi for all the encouragement, guidance, and support he has given me. It was an invaluable experience for me to be a member of his research team, and I felt supported throughout my studies. It would not have been possible for me to progress in my research without his constructive comments, insightful feedback, and endless support.

I would like to express my gratitude and appreciation to the faculty and the Computer Science program members.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CNN Convolutional Neural Network.

CSR Convolutional Sparse Representation.

DCT Discrete Cosine Transform.

DWT Discrete Wavelet Transform.

HIS Hue Intensity Saturation.

HS Hypers-Pectral.

HVS Human Visual System.

IF Image Fusion.

IQA Image Quality Assessment.

IR Infrared.

KHWT Kekre's Hybrid Wavelet Transform.

KWT Kekre's Wavelet Transform.

MAD Most Apparent Distortion.

MEF Multi-exposure Image Fusion.

MOS Mean Opinion Score.

MRI Magnetic Resonance Imaging.

MS Multi-Spectral.

MSE Mean Square Error.

MSSIM Mean Structural Similarity Index Measure.

NIR Near-Infrared.

NPS Noise Power Spectrum.

PCA Principal Component Analysis.

PDF Probability Density Function.

PET Positron Emission Tomography.

PSNR Peak Signal-to-Noise Ratio.

RGB Red–Green–Blue.

SAE Stacked Autoencoder.

SSIM Structural Similarity Index Measure.

SWT Stationary Wavelet Transform.

VIF Visual Information Fidelity.

VSNR Visual Signal-to-Noise Ratio.

Chapter 1

Introduction

1.1 Overview

Due to hardware device and optical imaging limitations, images captured by a single sensor or using a single shooting setting cannot fully depict the imaging scene. For instance, reflected lighting with limited brightness range and predefined depth of field represents incomplete information. Naturally, the fusion of images is able to combine meaningful information in different images to create a single, enhanced image that contains richer information and can be used in future applications [86]. The aim of image fusion is to extract information from input images and combine information to create a fused image that provides a better description of the scene for both human and machine perception compared to each individual input image. [39].

Image fusion is vital for image-based applications such as computer vision, robotics, remote sensing, and biomedical imaging. [52,64,76]. In photography, cameras are not able to capture an image with all focused objects due to limited depth of field. Objects that lie in the depth-of-field range are typically sharp in a camera-captured image, while the remaining objects tend to be blurred. In order to address this issue, multi-focus image fusion algorithms are proposed, which combine additional information from multiple im-

ages of a scene in order to create a fused image with an extended depth of field [29].

Medical image fusion involves a wide array of techniques that encompass both image fusion and general information fusion. These techniques are utilized to tackle medical issues determined by images of the human body, organs, and cells. As a result of the development of imaging technology, there has been an increasing interest in the application of these technologies in medical diagnostics, analyses, and historical documentation. Computer-aided imaging techniques facilitate the quantitative evaluation of images, thereby enhancing the effectiveness of medical practitioners in making unbiased and objective decisions within a short time frame. Furthermore, the use of multi-sensor and multi-source image fusion techniques expands the range of features used in medical analysis applications. This often results in robust information processing that is able to unveil hidden information that may not be visible to the human eye. The additional information derived from the fused images can be effectively utilized to enhance the precision of abnormality localization [31].

Visible images typically exhibit finer details compared to their corresponding infrared images. Infrared images often depict objects with higher contrast and intensity values, particularly when those objects have higher temperatures than their surroundings (thermal information captured in infrared images tends to manifest as coarser structures or features). Furthermore, infrared and visible images of the same scene can exhibit varying brightness responses and contrast values, resulting in substantial visual disparities. These disparities pose challenges for the observation and tracking of focal targets across both image types. However, infrared-visible image fusion technology offers a promising algorithmic approach to alleviate or compensate for this limitation. The fused image contains both infrared and visible information [85]. These applications emphasize the importance of image fusion techniques.

Methods of combining images can be divided into three categories: pixel-level, feature-level, and decision-level. Pixel-level image fusion involves combining the information

directly from the source images or from their multi-resolution transforms in order to produce a more informative final image. A feature-level fusion attempts to extract salient characteristics from the source images, such as shape, length, edges, segments, and directions. Feature extraction from input images is combined to create more meaningful features that provide a more comprehensive and descriptive image. A decision-level fusion represents a high level of fusion that provides an indication of the actual target. An ultimate fusion decision is produced by joining the results from several algorithms [70]. In this thesis, the focus is on pixel-level fusion techniques.

1.2 Problem Definition and Challenges

Most images have unilateral or limited information content. For example, a single sensor alone is often unable to provide a good-quality image in real-world applications where multiple optical sensors are used to acquire images. The same condition holds true in many modern technology fields, such as medicine. The term "good quality" itself refers to a wide range of aspects of an image scene, including illumination, sharpness, noise, and contrast [58]. There are also different types of multimodal images that convey different types of information. For instance, a variety of sensor tools are available, including optical cameras, millimeter wave cameras, infrared (IR) and near-infrared (NIR), X-rays, radar, and magnetic resonance imaging (MRI), all of which tend to emphasize a different aspect of a captured image. Besides sensor modalities, many other factors can also affect an image such as the occlusion of objects of interest due to smoke, fog, and other unwanted objects, changing illumination in scenery for photography applications (e.g., daylight exposure at different times of day), and adjustable parameters within the sensors themselves, such as focal length [58]. Notwithstanding the above, the use of multiple sensors is becoming increasingly popular for improving the capabilities of intelligent machines and systems. This has led to an intense focus of research and development on

multisensor fusion in recent years [42, 44, 48, 63, 75, 87].

The term multisensor fusion refers to the synergistic combination of different sources of sensory information into one representational format. The information to be combined may come from multiple sensory devices monitored over a specified time period, or from a single sensory device monitored over a longer period of time [49].

Using complementary information from disparate sensors with different modalities can enhance the performance of visualization, detection, and classification systems. Data from multiple sensors often provide complementary information about a scene or object, and thus image fusion is an effective method for comparing and analyzing these data. Multisensor image fusion offers several benefits, including wider spatial and temporal coverage, an extended range of operations, decreased uncertainty, increased reliability, and improved robustness [10].

Image fusion has gained significant popularity and found extensive applications in diverse fields, including medical diagnosis, surveillance, and photography. However, there are still various issues and challenges associated with image fusion in these domains [43]. The medical image fusion field presents several significant challenges. The first challenge is the absence of clinical problem-oriented fusion methods. The primary goal of medical image fusion is to enhance clinical outcomes. However, developing methods that specifically address a particular clinical problem remains a complex and difficult task, as it requires a combination of medical domain expertise and algorithmic insights. Another major challenge in medical image fusion is determining how to objectively assess fusion performance. The surveillance image fusion field encounters several major challenges, including computing efficiency and imperfect environmental conditions. In surveillance applications, effective image fusion algorithms should efficiently combine information from the original images to create a clear fused image. Moreover, surveillance scenarios often require continuous real-time monitoring. Therefore, an important aspect of surveillance applications is enhancing the speed of image fusion algorithms. Another significant

problem in the surveillance image fusion field is the presence of imperfect environmental conditions during image acquisition. For example, the source images may suffer from severe noise, underexposure caused by adverse weather conditions, or insufficient illumination. Developing image fusion methods that are robust to such imperfect conditions constitutes a crucial research area in this field. In the field of photography, image fusion faces significant challenges, particularly concerning the influence of moving targets and applications in consumer electronics. In photography applications, multi-focus and multi-exposure images are typically captured at different times. As a result, moving objects may appear at varying locations during the image capturing process. This presents a challenge for image fusion since moving objects can create ghost artifacts in the fused image. Moreover, in the realm of photography, the imaging process often involves capturing multiple shots with different camera settings, which can be time-consuming.

Image fusion is sometimes an introduction to another task. In this case, the effectiveness of fusion algorithms can be measured by the improvement in the following tasks. As an example, the number of correct classifications is the common measure of the evaluation of classification systems. For the purpose of this system evaluation, it is necessary to know the "true" correct classifications. However, the ground-truth data might not be available in experimental situations. It is crucial for a robust fusion algorithm to be guided by rigorous principles that ensure its capability to preserve the right amount of information from all images. Also, the human perception of the combined image is critical to many applications, which is why many fusion results are evaluated subjectively [10].

The Structural Similarity Index Measure (SSIM), originally proposed by Wang et al. [77, 88] can be considered a suitable model for the Human Visual System (HVS). The SSIM index assumes that the HVS has evolved to perceive visual distortions as changes in structural information. In practice, the SSIM is calculated locally. An 8×8 window moves, a single pixel at a time, across an image. A local SSIM score is calculated at each step. According to [81] the Mean SSIM (MSSIM) is the arithmetic average of all

local SSIM scores for an entire image. The SSIM has been widely accepted as one of the suitable measures of visual quality and, perceptual proximity between images based on subjective quality assessments involving large databases. With regard to the exceptional characteristics of the SSIM mentioned above, this thesis proposes using it to formulate image fusion as an inverse problem. The SSIM Mean definition, provided by Brunet in his thesis [5], is used as the data fidelity term in the inverse problem formulation. Gradient-based methods are used to solve this inverse problem numerically. However, it is crucial to acknowledge that this study may possess certain limitations. The selection of the starting point and parameters for the gradient-based methods significantly influence the numerical solution process. Additionally, a primary challenge in this research refers to calculating the derivative of the SSIM Mean, as the problem is solved using gradient-based methods. It should be noted that Mean SSIM and SSIM Mean are distinct measures. Mean SSIM refers to the averaging of SSIM values calculated over a bunch of local windows or patches. On the other hand, SSIM Mean introduces a novel approach to averaging by using SSIM.

In this thesis, SSIM's potential is demonstrated as a perceptual quality measure that will be used to optimize primary image fusion algorithms. A framework for combining images is presented based on optimizing an objective quality measure formed using the SSIM.

1.3   Contribution

A general mathematical and computational framework is provided for image fusion, utilizing SSIM Mean as the "data fitting" term in the objective function of the inverse problem.

It is common to formulate the recovery of images as an inverse problem with a linear combination of data consistency error and regularization penalty as a criterion. The regularization function penalizes images that do not exhibit desired properties (such as

piecewise smoothness and sparsity). Here the basic idea is that replacing the widely used

Euclidean distance, which is known as a poor measure of visual quality [59, 60, 77, 88],

with the SSIM, which is considered to be one of the best measures of visual proximity

would improve the quality of visual experience. To the best of our knowledge, none of

the previous works have attempted to solve it in this form [60].

The structural similarity index measure is defined as a function of mean, variance, and

covariance. Another expression of SSIM is subsequently introduced, which employs the

mean and vector norm function instead of mean, variance, and covariance. This alterna-

tive expression offers easier mathematical calculations and is applied to the SSIM Mean

definition provided by Brunet in his work [5]. However, it should be noted that our

approach is different from the appraoch in [5]. SSIM Mean is a weighted average scheme

that aims to find an ideal image/vector that maximizes the sum of the SSIM values of

a set of input images/vectors. The SSIM Mean would be the suitable choice for fusion

images due to SSIM's characteristics since fusion image methods are intended to create

a fused image that combines information from multiple input images, resulting in an

improved representation of the scene compared to any of the individual images alone.

The following problem is intended to be solved: Given $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ a set of n vectors

with the size of $\mathtt{n} \times 1$ and $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ a set of associated weights, $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_n\}$ a

set of degradation operators with the size of $\mathtt{n} \times \mathtt{m}$, $g(\mathbf{x})$ is a regularization term, and λ is

its corresponding regularization parameter, find $\mathbf{x}$ with the size of $\mathtt{m} \times 1$ which maximizes

$$\sum_{i=1}^{n} \mathbf{p}_i \, \mathrm{SSIM}(\mathcal{H}_i\mathbf{x}, \mathbf{y}_i) + \lambda g(\mathbf{x}). \tag{1.1}$$

The above problem is solved numerically and in an iterative manner in this thesis. As

gradient-based methods are employed to solve this problem, the derivative of SSIM-Mean

is computed. It should be noted that this approach differs from Brunet's approach.

While he calculates only the SSIM Mean for a set of vectors and obtains a closed-form

solution without flexibility for incorporating regularization, the major advantage of the

formulation 1.1 is its flexibility, i.e. the incorporation of additional terms as well as the

use of different types of regularization terms. Furthermore, the equation 1.1 provides a formulation that can be used to solve a wide range of inverse problems due to its flexibility to apply any arbitrary linear degradation operators $\mathcal{H}_i$, although note that Brunet's formulation only considers identity matrix as $\mathcal{H}_i$.

SSIM Mean derivatives are mostly theoretical and complementary to Brunet's approach. It is shown that the SIMM Mean for a set of points and vectors can be computed using numerical methods and obtain the same result as Brunet's.

The piecewise method is also discussed, which involves the use of MSSIM instead of SSIM. The statistics of images vary widely across their spatial domain and it leads to the MSSIM, which is an averaging of local SSIM measurements, being used to compute the similarity between two images [61, 88].

Assuming the same variable and parameter definitions as equation 1.1, the problem is to find $\mathbf{x}$ which maximizes

$$\sum_{i==1}^{n} \mathbf{p}_i \ \mathrm{MSSIM}(\mathcal{H}_i\mathbf{x}, \mathbf{y}_i) + \lambda g(\mathbf{x}). \tag{1.2}$$

To solve it numerically using a gradient-based method, the derivative of MSSIM, as provided in [59], is applied in the numerical method.

According to our knowledge, all of the previous works for solving 1.2 are performed by assumption $\mathcal{H}_i = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

1.4   Thesis Outline

This thesis seeks to provide a framework for fused images based on optimization methods by incorporating a perceptual criterion, SSIM index. In order to improve image processing algorithms, the perceptual criterion must be taken into consideration. Indeed, algorithms have to be optimized according to a perceptually relevant model in order to obtain better image quality.

In this thesis, the chapters are divided as follows.

- Chapter 2 introduces the necessary definitions and concepts related to inverse problems, regularization methods, image quality, and the notion of SSIM. As part of this Chapter, As part of this chapter, SSIM is presented as a function of mean and vector norm.

- Chapter 3 presents existing approaches for visualization and understanding of image fusion approaches.

- In Chapter 4, The proposed formulation is presented for image fusion as an inverse problem, utilizing SSIM Mean as the data fidelity term. The problem is solved numerically by the gradient-based method.

- Chapter 5 presents the results of the proposed framework, including a comparison of the SSIM Mean between the proposed method and Brunet's work [5]. The results of the proposed method on images are examined quantitatively and qualitatively using a comprehensive set of examples.

- In Chapter 6, The main contributions of the thesis research are discussed, along with some limitations. Furthermore, suggestions for interesting future directions of the thesis research are provided.

1.5 Software & Source Code

Matlab and Python are used to implement the proposed models. Python is initially used for implementing the proposed model. However, since Brunet's code provided in his thesis is written in Matlab, it is used for later experiments in this study. Python programming language is an interpreted, general-purpose, high-level language. Its design philosophy emphasizes code readability with the use of significant indentation. It is designed to be an easy-to-read language. The following Python packages were used for this thesis.

- NumPy is a Python package created for the purpose of scientific computing in Python. The program is capable of handling large, multi-dimensional arrays and matrices, along with several mathematical operations and functions that can be applied to these arrays. https://numpy.org/

- Scikit-image, or skimage, is an open source Python package designed for image preprocessing, and computer vision. https://scikit-image.org/

The MATLAB language is a high-performance language used for technical computing. This program integrates computation, visualization, and programming in a familiar mathematical notation that makes it easy for users to understand.

Source Code

You can access the Matlab and Python implementations of our models, evaluation metrics, and visualization of the explanations at the following link.

`https://github.com/fatemehhirbodvash/SSIM-Mean-.git`

Chapter 2

Background

2.1 Imaging and Inverse Problems

2.1.1 General Model of an Inverse Problem

The features of most interest in the field of applied sciences cannot be observed directly, but have to be inferred from other observable quantities. In the simplest case, which is valid in a wide range of situations, there is a linear relationship between the features of interest and the derived quantities. Assuming that the object is modeled by a function x and the observed quantities or images by another function $\mathbf{h}$, we can formulate the problem of inferring $\mathbf{x}$ from $\mathbf{h}$ as an inverse problem, which consists of solving an equation,

$$\mathcal{H}(\mathbf{x}) = \mathbf{h}, \tag{2.1}$$

where $\mathcal{H}$ is a (typically linear) operator.

In the real world, the observations or data we model with $\mathbf{u}$ are not exactly equal to the image $\mathcal{H}(\mathbf{x}) = h$, but rather to a distortion of $\mathbf{h}$. The distortion is commonly modeled using an additive noise or error term $\mathbf{n}$, i.e.

$$\mathbf{u} = \mathbf{h} + \mathbf{n} = \mathcal{H}(\mathbf{x}) + \mathbf{n}. \tag{2.2}$$
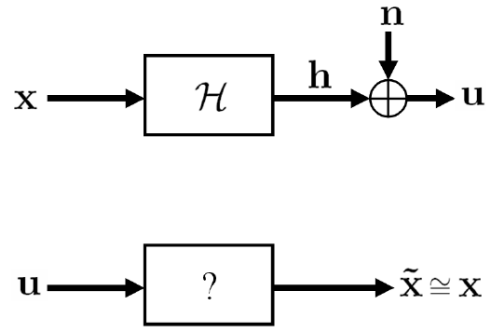
Figure 2.1: Inverse problem [16].

Finally, the inverse problem can be summarized in the following statement: Using the observations $\mathbf{u}$, find an approximation of the object $\mathbf{x}$ denoted by $\tilde{\mathbf{x}}$ (see Figure 2.1) [16].

2.1.2 Ill-Posed Problems

From the point of view of modern mathematics, all problems can be classified as being either well-posed or ill-posed. Consider the operator equation

$$\mathcal{H}(\mathbf{x}) + \mathbf{n} = \mathbf{u}, \quad \mathbf{x} \in X, \quad \mathbf{u} \in U \tag{2.3}$$

where $\mathbf{X}$ and $\mathbf{U}$ are metric spaces [16]. According to Jacques Hadamard [25], the problem 2.3 is said to be well-posed if the following two conditions hold [72],

1. For each $\mathbf{u} \in U$ the equation 2.3 has a unique solution;

2. The solution of 2.3 is stable under perturbation of the righthand side of this equation, i.e. the operator $\mathcal{H}^{-1}$ is defined on all of U and is continuous.

Many image processing problems are ill-posed in the sense of Hadamard, such as denoising, deblurring, inpainting, and so on [21].

2.2   Noise Model

Noise is a random signal. It destroys most of the image information. Digital images can be distorted by different types of noise including Gaussian noise, Poisson noise, Speckle noise, Salt and Pepper noise, and many other types of noise. These noises may be derived from a noise source present in the vicinity of image capturing devices, faulty memory location, or may be introduced due to imperfection or inaccuracy in image capturing devices such as cameras, misaligned lenses, weak focal length, scattering, and other adverse conditions present in the atmosphere [4].

2.3   Probability Functions for Discrete Random Variables

The various noises considered here are random in nature. Their exact values are random variables whose values are best described using probabilistic notions. Since their exact values are random variables, probabilistic notions are the best means of describing their exact values.

Let Z denote a random variable with values z and range $\mathbb{R}$ where is a real number. A probability function, $p(z)$, lists (literally or functionally) the probabilities of all possible values of z, and satisfies the following conditions

$$1)\, p(z) \geq 0, \quad \text{for all } z \in \mathbb{R}$$
$$2)\, \sum_{z \in Z} p(z) = 1 \tag{2.4}$$

where $p(z)$ means $P(Z = z)$. $P(Z = z)$ is the probability that random variable Z has the specific value z from $\mathbb{R}$ which is real number [22]. In the case of continuous random variables, $P(Z = z)$ is known as a Probability Density Function (PDF). In the context of discrete quantities that approximate continuous quantities (for example, image intensity),

we typically use the term PDF to emphasize that the underlying random variables are continuous [22].

2.3.1   Gaussian Noise

Many natural sources of Gaussian noise exist, such as the thermal vibrations of atoms in an antenna [13]. Additive Gaussian noise models are widely used in practice due to their mathematical tractability in both the spatial and frequency domains.

The probability density function (PDF) of a Gaussian random variable, z defines as

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\bar{z})^2}{2\sigma^2}} \qquad (2.5)$$

where z represents intensity, $\bar{z}$ is the mean (average) value of z, σ is its standard deviation, and e is exponential function [22].

2.3.2   White Noise

This is noise with a constant power spectrum, i.e. its spectral density, commonly known as the Noise Power Spectrum (NPS), is constant with frequency. Theoretically, the spectrum would extend to infinite frequency and therefore the total noise power would be infinite; in practice, the spectrum of any naturally occurring white noise falls off at sufficiently high frequencies. The term derives from the analogy of white light, which contains nearly all the frequencies in the visible spectrum equally [13, 22].

White noise is completely uncorrelated, i.e. the value of each pixel is independent of its neighboring pixels. Thus, its autocorrelation function is zero. However, being uncorrelated does not limit the range of possible values that a signal can take; it is possible for the signal to have any distribution of values. For example, a binary signal that can only take on the values of 1 or 0 is considered "white" when the sequences of ones and zeros are statistically uncorrelated. There is also the possibility of white noise having a continuous distribution, such as a normal distribution [13].

2.3.3 Salt and Pepper Noise

Another common form of noise is data drop-out noise, commonly referred to as salt-and-pepper noise or impulse noise. Data transmission errors cause this noise. The images are distorted by corrupted pixels, which are either set to zero or to the maximum value. These results produce a "salt and pepper" appearance to the images. Unaffected pixels remain unchanged. The noise is usually quantified by the percentage of pixels which are corrupted [13].

A more detailed description of the types of noises can be found in the standard textbook on "Digital Image Processing" [22]. The next section introduces zooming as a degradation model that would be used later on in this thesis.

2.4 Zooming

We consider the following problem of image zooming. Given the observed image data y, where

$$y = \mathcal{D}_z(x) + \mathbf{n}, \tag{2.6}$$

in which $\mathcal{D}_z(x)$ is the downsampling by a factor of z, find an approximation of x.

Note that we defined the local averaging operator of length $\mathbf{z}$, $\mathcal{A}_z : \mathbb{R}^{M \times N} \longmapsto \mathbb{R}^{M \times N}$ where and $M \times N$ is the size of the image and $\mathbb{R}$ is real number (i.e. $\mathbb{R}^{M \times N}$ is a set of all vectors which their size is $M \times N$ with real-valued elements.) for any $(i, j) \in [1, \cdots, M] \times [1, \cdots, N]$,

$$\mathcal{A}_z x(i, j) = \frac{1}{z^2} \sum_{0 \leq i' < z, 0 \leq j' < z} x(i + i', j + j'). \tag{2.7}$$

Also, the downsampling operator $\mathcal{S}_z : \mathbb{R}^{M \times N} \longmapsto \mathbb{R}^{M \times N}$ is defined for any image $x \in \mathbb{R}^{M \times N}$ such that for any $(i, j) \in [1, \cdots, M] \times [1, \cdots, N]$,

$$\mathcal{S}_z x(i, j) = x((i - 1)z + 1, (j - 1)z + 1). \tag{2.8}$$

Note that $\mathcal{D}_z = \mathcal{S}_z \circ \mathcal{A}_z$.

2.4.1 Regularization-based Method

In order to solve ill-posed problems, one must overcome the instability estimating algorithms of their solutions. To mitigate this instability, theoretical strategies and numerical tools have been developed, and regularization theory has become mainstream. In regularization, the ill-posed problem pointing to an unstable solution is replaced with a well-posed problem pointing to an approximate solution based continuously on the given data [20]. The purpose of the regularization is to prevent over-fitting by introducing *prior* information about problems to penalize the model. The *prior* information is usually a penalty for complexity, such as restrictions for smoothness in total variational regularization [65] and bounds on the norm of the space vector in Tikhonov regularization [71,73]. This term is usually known as a penalty term.

The general regularization model can be formulated as

$$f(\mathbf{x}) + \lambda g(\mathbf{x}), \tag{2.9}$$

where $f(\mathbf{x})$ is known as fidelity term and here is the function defined in the equation 2.2, $g(\mathbf{x})$ is regularization term or penalty term, and λ is regularization parameter.

The regularization parameters are introduced naturally as a way to adjust the dominator between the loss and penalty term in order to achieve a better compromise [20]. As the regularization parameter approaches zero, the functions of the family tend to converge to the exact solution of the problem for noise-free data. In the case of noisy data, With a non-zero regularization parameter, an optimal approximation of the exact solution can be achieved [50].

2.4.2   Image Quality Assessment (IQA)

The use of digital images as a means of presenting and communicating information has grown rapidly in the early 21st century [82]. According to the Domo `https://www.domo.com/blog/data-never-sleeps-hits-double-digits/` data Never Sleeps in 2022 turns over ten years. Based on an analysis of infographics during these years, it has become apparent that the largest increases are associated with content sharing via images and videos. There is no doubt that this makes sense with the growing spread of smartphones with better cameras, WiFi coverage, and cell data plans, which facilitate data transmission. However, digital images go through a series of processing steps that can introduce various distortions and lowering perceptual quality before reaching the end user. It is, therefore, crucial to have a reliable tool for assessing, controlling, and ensuring quality [14], i.e. image acquisition, management, communication, and processing systems should be able to identify and quantify image quality degradations in order to maintain, control, and enhance the quality of images. In order to achieve this objective, it is essential to develop effective automated systems for assessing image quality [82].

The measurement of image quality is an integral part of the vast majority of image processing and fusion applications. For example, measuring the quality of images produced by medical devices when designers want to determine which devices produce the best results [18]. It is possible to use quality assessment metrics as testing criteria or as optimization goals for visual communication systems. Methods for assessing quality can be categorized into subjective and objective approaches.

2.4.2.1   Subjective Image Quality Assessment

In most visual communication systems, the human visual system is the ultimate receiver of visual signals. Therefore, subjective assessment is generally considered the most accurate and reliable method. In this method, a group of people is asked to give their opinion about the quality of each image in a given data set. Here are a few standardized

subjective IQA methods that will be briefly introduced in the following [56].

- **Single Stimulus Categorical Rating**

  In this approach, test images are displayed on a screen for a predetermined duration. Subsequently, the images disappear from the screen, and observers are requested to evaluate their quality using an abstract scale consisting of the categories: excellent, good, fair, poor, or bad. The display of test images is randomized. There are some methods that use continuous scales rather than categorical scales in order to avoid quantization artifacts.

- **Double Stimulus Categorical Rating**

  This method is similar to the single stimulus method. This method, however, displays both the reference and test images for a fixed duration. As soon as the images have disappeared from the screen, observers are asked to rate the quality of the test image using the abstract scale described earlier.

- **Ordering By Force-choice Pair-wise Comparison**

  Two images of the same scene are displayed to observers in this type of subjective assessment. Then, they are asked to select the image that has the highest quality. Even if there is no difference between the two images, observers are required to select one image. Observers may make their decision at any time.

- **Pair-wise Similarity Judgments**

  It has already been mentioned, in force-choice comparisons, observers are required to select one image regardless of whether they perceive any difference between the two images. In pair-wise similarity judgments, observers must not only choose the image with higher quality but also indicate the level of difference between them.

Using raw rating results such as excellent, good, fair, and so on as quality scores might seem tempting. However, these ratings are unreliable. One of the reasons is that observers tend to assign different quality scales to each scene and even different types of

distortions. Two scoring methods, Difference Mean Opinion Score (DMOS) and Z-score are commonly employed in subjective IQA. Modern IQA metrics use differences in quality between images instead of directly applying rating results. DMOS is measured by the difference between the raw quality score of the reference and test images. To facilitate the comparison of individual observers' opinions on image quality, a linear transformation is applied that makes the mean and variance equal for all observers. The result of this transformation is referred to as the Z-score [56].

Subjective evaluation of image quality can be an accurate and reliable way to measure image quality. However, the subjective test is time-consuming and expensive, and cannot be directly embedded into a practical system as the optimization metric [14, 79]. Furthermore, subjective experiments face challenges due to various factors including viewing distance, display device, lighting conditions, subjects' visual abilities, and subjects' moods. Therefore, the design of mathematical models capable of accurately predicting the quality assessment of an average human observer becomes essential [56].

2.4.2.2 Objective Image Quality Assessment

To overcome the limitations of subjective image quality assessment, objective image quality assessment aims to construct computational models that predict the perceived quality of visual images, i.e. the objective quality metrics are mathematical models that calculate the image quality. Objective image quality metrics are generally classified based on their dependence on the availability of an original reference image. Distorted images are compared to a complete reference image in the full-reference method. If partial information about the reference image is required, it is referred to as a reduced reference method, while no-reference (or blind) methods are limited to the distorted image [36].

2.5 Applications of IQA

In general, image quality metrics have three main applications [88].

1. It can be set up as a quality control system to monitor image quality. For example, an image and video acquisition system can automatically adjust itself based on the quality metric to obtain the best image and video data.

2. It can be used to measure the performance of image processing systems and algorithms. Suppose it is necessary to choose one of several image processing systems for a specific task. In that case, a quality metric can help to determine which one produces the highest-quality images.

3. It can be integrated into an image processing system to optimize the algorithms and parameter settings. For example, a quality metric in a visual communication system enables to help with optimal design of the prefiltering and bit assignment algorithms at the encoder and the postprocessing algorithms at the decoder.

Image denoising [22], deblurring [22, 24], and inpainting [23, 30] are only a few examples of standard image-processing tasks which are traditionally solved through numerical optimization. In most cases, the objective function associated with such problems is expressed as the sum of a data fidelity term f and a regularization term g (or a number thereof). In particular, considering the desired image estimate $\mathbf{x}$ to be a (column) vector in $\mathbb{R}^n$, both f and g are usually defined as non-negative functionals on $\mathbb{R}^n$, in which case the standard form of an optimization-based imaging task is given by

$$\min_{\mathbf{x}} \quad f(\mathbf{x}) + \lambda g(\mathbf{x}) \tag{2.10}$$

Here $\lambda \geq 0$ is a regularization constant that balances the effects of empirical and prior information on the optimal solution [5].

2.6 Mean Square Error (MSE)

A signal fidelity measurement compares two signals by giving a numerical score describing the degree of similarity/fidelity or, conversely, the level of error or distortion between them. It is usually assumed that one of the signals is an original signal while the other is distorted or contaminated with errors [78]. One of the most widely used methods for measuring image fidelity is MSE.

The MSE between two images y_1 and y_2 is

$$MSE(y_1, y_2) = \frac{1}{L_1 L_2} \sum_{i_1=1}^{L_1} \sum_{i_2=1}^{L_2} (y_1(i_1, i_2) - y_2(i_1, i_2))^2, \tag{2.11}$$

where L_1 and L_2 correspond to the length and width of the images, respectively. Since MSE is computed based on the error signal, $e = y_1 - y_2$, between the reference image, y_1, and its distorted version, y_2. It is common to convert MSE into peak signal-to-noise ratio(PSNR) using the following expression in image processing

$$PSNR(y_1, y_2) = 10 \log_{10} \frac{R^2}{MSE(y_1, y_2)}, \tag{2.12}$$

where R is the dynamic range of image pixel intensities e.g. for an 8-bit/pixel gray-scale image, $R = 2^8 - 1 = 255$. In terms of perceptual quality, PSNR provides an advantage over MSE in that it can handle images with different dynamic ranges. MSE has been widely used in the literature as a signal fidelity measure and its use as an image/video quality assessment metric has become a convention.

The poor performance of MSE as a method for assessing image/video quality is over-looked in favor of its attractive features, such as simple operation, low computational costs, and memorylessness [82]. Due to the following reasons, MSE is well suited for solving optimization and design problems: it is a valid distance metric in $\mathbb{R}^n$; After any orthogonal (or unitary) linear transformation, energy is preserved (Parseval's theorem; This theorem states that the sum of the absolute squares of a function defined at discrete, equally spaced points is equal to the sum of the absolute squares of its Fourier

(a)   (b)   (c)

(d)   (e)   (f)

Figure 2.2: Comparison of distorted images with the same MSE. (a) Original image (MSE=0;SSIM=1); (b) Shift (to the left) (MSE=400;SSIM=0.97); (c) Rotation (clockwise) (MSE=430.24;SSIM=0.67); (d) Blurring (MSE=402;SSIM=0.62); (e) Gaussian noise (MSE=400;SSIM=0.34); (f) Salt and pepper noise (MSE=402;SSIM=0.57);

coefficients [41].); it is convex and differentiable; Closed-form or iterative numerical solutions are often available for optimization problems; it is additive for independent sources of distortions. Although MSE has solid mathematical properties, it should not be used unquestioned as a perceptual quality measure in image and video processing.

Figure 2.2 illustrates and rationalizes why it is not reasonable to rely on MSE's judgment of perceptual quality. A reference image is shown in Figure 2.2 (a). The rest of the images are generated using the reference image by introducing the same level of distortion in terms of MSE to the reference image. It is easily seen that the perceptual

quality of the distorted images differs significantly, despite MSE incorrectly predicting a similar quality. In Figures 2.2 (b) to (f), the MSE indicates that the distortion levels are approximately similar. In contrast, the image in Figures 2.2(b) shows the least distortion perceptually, according to the HVS. Several psychological and physiological features of the HVS are not accounted for by the MSE [82].

The reason why MSE is unable to provide accurate perceived quality predictions lies in a number of questionable assumptions when used as an image/video quality measure. 1) As far as perceptual quality is concerned, the spatial relationships between pixels are irrelevant; therefore, distortion in each pixel can be determined individually; 2) In any reference image, the error signal, e, introduces the same level of distortion; 3) The sign of e does not affect the evaluation of perceptual quality ; 4) All the pixels of an image are equally important for perceptual image quality. All of these implicit assumptions are extremely strong since they impose significant limitations on the signal samples, their interactions with each other, and their interactions with errors. In terms of measuring visual perception of image fidelity, none of them hold (even roughly) [78].

2.7   Structural Similarity

It is the classical approach for IQA to attempt to model every single component of the HVS. This technique is referred to as the bottom-up approach. It means that we start with simple features of the HVS and combine them until we obtain a model of image quality perception [5]. As an alternative to the bottom-up approach, Wang et al [88] proposed the structural similarity measure index. The SSIM provides a measure of visual closeness between an image and a distorted or corrupted version of it. Due to the assumption that a distortionless image is always available, the SSIM can be considered as a full-reference measure of image quality assessment [88].

This definition is based on two assumptions: (i) images are highly structured, i.e. pixels
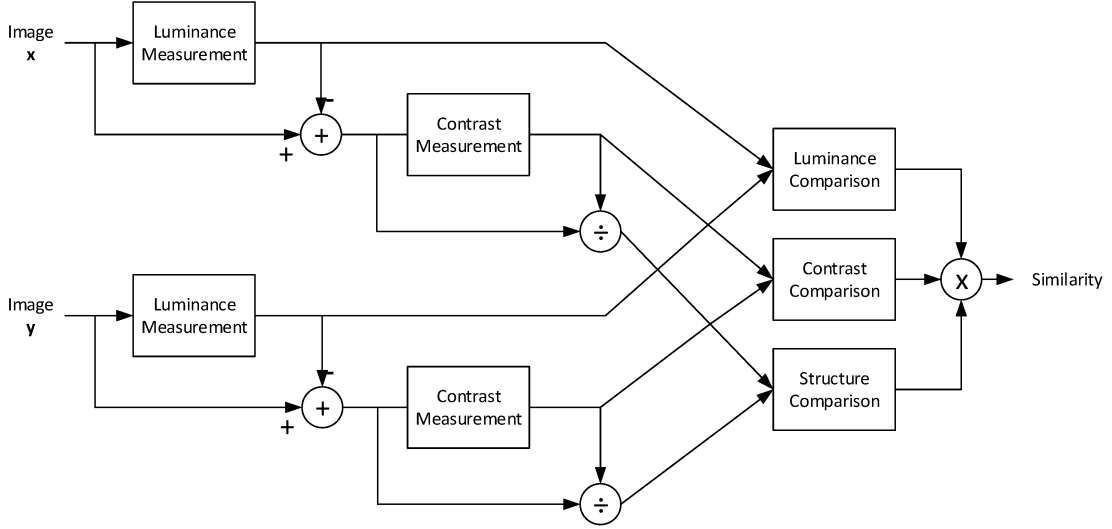
Figure 2.3: Diagram of the structural similarity (SSIM) measurement system [88].

tend to be correlated, especially if they are spatially close, and (ii) the HVS is capable of extracting structural information. This is why SSIM quantifies changes in perceived structural information to measure similarity. The system diagram of the proposed quality assessment system is shown in Figure 2.3.

By comparing the luminance, contrast, and structure of two images, this measurement can be performed. In order to quantify changes in luminance between two images, we quantify relative changes in their means given two images $x \in \mathbb{R}^{n \times m}$ and $y \in \mathbb{R}^{n \times m}$ (i.e. x and y are $n \times m$ dimensional image.) [61]. In order to do this, we use the following formula

$$l(x, y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}. \tag{2.13}$$

The maximum value of this factor is one when $\mu_x = \mu_y$ [28]. In this case, it is important to emphasize that $l(x, y)$ is sensitive to the relative changes in luminance, not to its absolute change. Accordingly, this is in accordance with Weber's law, which is a model for light adaptation of the HVS [88]. According to Weber's law, the magnitude of a just-noticeable luminance change ΔI is approximately proportional to the background luminance I for a wide range of luminance values. In other words, the HVS is sensitive

to the relative rather than the absolute luminance change [80].

For the contrast, the relative variance is used to measure the comparison

$$c(x,y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \tag{2.14}$$

In order to avoid division by zero, $C_2 > 0$ is added. A contrast is maximal and equal to one if x equals y.

Calculating the correlation coefficient between images x and y compares the two structures simply

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}. \tag{2.15}$$

For stability, the positive constant C_3 is again used. The final step is to combine these three components using the function $f(x_1, x_2, x_3) = x_1^\alpha x_2^\beta x_3^\gamma$,

$$\text{SSIM}(x,y) = l(x,y)^\alpha c(x,y)^\beta s(x,y)^\gamma. \tag{2.16}$$

The relative influence of the three comparisons is controlled by positive parameters α, β, and γ. The authors simplify 2.16 by setting $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$ in [88].

$$\begin{aligned}
\text{SSIM}(x,y) &= \left(\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right)\left(\frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right)\left(\frac{2\sigma_{xy} + C_2}{2\sigma_x \sigma_y + C_2}\right), \\
&= \left(\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right)\left(\frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right).
\end{aligned} \tag{2.17}$$

Setting,

$$\text{SSIM}(x,y) = S_1(x,y)S_2(x,y), \tag{2.18}$$

where

$$S_1(x,y) = \left(\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}\right), \tag{2.19}$$

$$S_2(x,y) = \left(\frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}\right). \tag{2.20}$$

The authors in the original paper of SSIM chose [88] $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ where L is the dynamic range of the pixel values (255 for 8-bit grayscale images), and $K_1 \leq 1$ and $K_2 \leq 1$ are a small constants. This definition of SSIM will be used throughout

the remainder of this thesis. A positive SSIM index value is in the range [0,1]. A value of 0 means no correlation between images, and 1 means that $x = y$.

Figure 2.2 illustrates SSIM's ability to distinguish structural distortion from nonstructural distortion. It has been shown that shift (to the left) (b) causes nonstructural distortions to the reference image (Figure 2.2(a)) and is, therefore, less penalized by the SSIM index compared with structural distortions such as rotation(clock-wise) (Figure 2.2(c)) blurring (Figure 2.2(d)), Gaussian noise (Figure 2.2(e)), and salt and pepper noise (Figure 2.2(f)).

2.8   MSSIM

Local SSIM methods are used rather than global SSIM methods, using a sliding Gaussian window of fixed size to determine the SSIM for each point of an image. The result of this analysis is a map that indicates how image quality varies spatially. Calculating the average of SSIMs can yield a global SSIM. It is in this case that mean SSIM is introduced,

$$\text{MSSIM}(x, y) = \frac{1}{N} \sum_{i=1}^{N} \text{SSIM}(x_i, y_i). \tag{2.21}$$

The number of local windows in the image is N, and x_i and y_i represent the ith sample sliding square windows, also known as patches in images x and y.

2.9   Another Expression for SSIM

In the definitions of SSIM and MSSIM that have been presented until now, the assumption is that x, y are images. There is no doubt that this definition is satisfied for signals (vectors). A new expression of the SSIM formula is presented in this section in terms of the mean of vectors $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{y} \in \mathbb{R}^n$ and the use of norms instead of variance and covariance. As a result of this alternative expression, SSIM derivatives can be taken in a practical way.

By considering, $\mu_{\mathbf{x}} = \bar{\mathbf{x}}$, $\mu_{\mathbf{y}} = \bar{\mathbf{y}}$, and $C_1 = \varepsilon_1$ the luminance function 2.13 can be written as

$$S_1(\mathbf{x}, \mathbf{y}) = \frac{2\mu_{\bar{\mathbf{x}}}\mu_{\bar{\mathbf{y}}} + C_1}{\mu_{\bar{\mathbf{x}}}^2 + \mu_{\bar{\mathbf{y}}}^2 + C_1} = \frac{2\bar{\mathbf{x}}\bar{\mathbf{y}} + \varepsilon_1}{\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2 + \varepsilon_1} = S_1(\bar{\mathbf{x}}, \bar{\mathbf{y}}). \tag{2.22}$$

The properties of mean, variance, and covariance indicate that $\sigma_{\mathbf{x}-\bar{\mathbf{x}}} = \sigma_{\mathbf{x}}$, $\sigma_{\mathbf{y}-\bar{\mathbf{y}}} = \sigma_{\mathbf{y}}$, and $\sigma_{(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{y}-\bar{\mathbf{y}})} = \sigma_{\mathbf{xy}}$. Thus, we can represent 2.17 as follows

$$S_2(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_{\mathbf{xy}} + C_2}{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + C_2} = \frac{2\sigma_{(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{y}-\bar{\mathbf{y}})} + C_2}{\sigma_{\mathbf{x}-\bar{\mathbf{x}}}^2 + \sigma_{\mathbf{y}-\bar{\mathbf{y}}}^2 + C_2} = S_2(\mathbf{x} - \bar{\mathbf{x}}, \mathbf{y} - \bar{\mathbf{y}}). \tag{2.23}$$

As a result of defining mean, variance, and covariance, we have $\mu_{\mathbf{x}-\bar{\mathbf{x}}} = \mu_{\mathbf{x}} - \bar{\mathbf{x}} = \bar{\mathbf{x}} - \bar{\mathbf{x}} = 0$ and similarly, $\mu_{\mathbf{y}-\bar{\mathbf{y}}} = 0$, then,

$$\sigma_{\mathbf{x}-\bar{\mathbf{x}}}^2 = \frac{1}{n-1} \sum_{i=1}^{n} ((\mathbf{x} - \bar{\mathbf{x}}) - 0)^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x - \bar{\mathbf{x}})^2 = \frac{\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2}{n-1}, \tag{2.24}$$

$$\begin{aligned} \sigma_{(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{y}-\bar{\mathbf{y}})} &= \frac{1}{n-1} \sum_{i=1}^{n} ((\mathbf{x} - \bar{\mathbf{x}}) - 0)((\mathbf{y} - \bar{\mathbf{y}}) - 0) \\ &= \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x} - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}}) \\ &= \frac{(\mathbf{x} - \bar{\mathbf{x}})^T (\mathbf{y} - \bar{\mathbf{y}})}{n-1}. \end{aligned} \tag{2.25}$$

By substituting 2.24 and 2.25 in 2.23, it can be written

$$\begin{aligned} S_2(\mathbf{x} - \bar{\mathbf{x}}, \mathbf{y} - \bar{\mathbf{y}}) &= \frac{2\sigma_{(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{y}-\bar{\mathbf{y}})} + C_2}{\sigma_{\mathbf{x}-\bar{\mathbf{x}}}^2 + \sigma_{\mathbf{y}-\bar{\mathbf{y}}}^2 + C_2} \\ &= \frac{\frac{2(\mathbf{x}-\bar{\mathbf{x}})^T(\mathbf{y}-\bar{\mathbf{y}})}{n-1} + C_2}{\frac{\|\mathbf{x}-\bar{\mathbf{x}}\|_2^2}{n-1} + \frac{\|\mathbf{y}-\bar{\mathbf{y}}\|_2^2}{n-1} + C_2}, \end{aligned} \tag{2.26}$$

and $\varepsilon_2 = (n-1)C_2$,

$$\begin{aligned} S_2(\mathbf{x} - \bar{\mathbf{x}}, \mathbf{y} - \bar{\mathbf{y}}) &= \frac{2(\mathbf{x} - \bar{\mathbf{x}})^T (\mathbf{y} - \bar{\mathbf{y}}) + \varepsilon_2}{\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 + \|\mathbf{y} - \bar{\mathbf{y}}\|_2^2 + \varepsilon_2} \\ &= \frac{\|\mathbf{x} - \bar{\mathbf{x}} + \mathbf{y} - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 - \|\mathbf{y} - \bar{\mathbf{y}}\|_2^2 + \varepsilon_2}{\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 + \|\mathbf{y} - \bar{\mathbf{y}}\|_2^2 + \varepsilon_2}. \end{aligned} \tag{2.27}$$

Therefor the SSIM formula 2.17 can be rewritten as,

$$\mathrm{SSIM}(\mathbf{x}, \mathbf{y}) = S_1(\bar{\mathbf{x}}, \bar{\mathbf{y}}) S_2(\mathbf{x} - \bar{\mathbf{x}}, \mathbf{y} - \bar{\mathbf{y}}), \tag{2.28}$$

$$\mathrm{SSIM}(\mathbf{x}, \mathbf{y}) = \Big(\frac{2\bar{\mathbf{x}}\bar{\mathbf{y}} + \varepsilon_1}{\bar{\mathbf{x}}^2 + \bar{\mathbf{y}}^2 + \varepsilon_1}\Big)\Big(\frac{\|\mathbf{x} - \bar{\mathbf{x}} + y - \bar{\mathbf{y}}\|_2^2 - \|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 - \|\mathbf{y} - \bar{\mathbf{y}}\|_2^2 + \varepsilon_2}{\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 + \|\mathbf{y} - \bar{\mathbf{y}}\|_2^2 + \varepsilon_2}\Big). \tag{2.29}$$

Chapter 3

Literature Review

Recent research has focused on developing objective quality assessment measures that can measure perceived distortions in visual content. Some of the more prominent are the structural similarity index [88] and derivatives [59, 83], the VVisual Information Fidelity (VIF) [68], the Visual Signal-to-Noise Ratio (VSNR) [9], and the Most Apparent Distortion (MAD) [40]. Compared with other methods, SSIM is often preferred because of its good trade-off between accuracy, simplicity, and efficiency [78]. Here are some examples of SSIM-based imaging tasks that can be found in the literature. For instance, in [1] the authors use SSIM to measure quality, particularly for sparse representation, whereas previous studies used MSE. To solve a global optimization problem in maximum SSIM, gradient descent is used. In this work, they were focusing on the improvement of image restoration algorithms with regard to visual quality, which takes sparsity *prior* on the solution based on a dictionary. Brunet developed some necessary tools for the design of SSIM-optimal algorithms in his thesis [5]. In this work, theoretical developments are combined with experimental research and practical algorithms. The mathematical properties of the SSIM index are described. It is demonstrated how SSIM indexes can indeed be transformed into distance metrics. There are also proofs of local convexity, quasi-convexity, symmetries, and invariance properties. A number of analytical techniques are

then developed for different kinds of SSIM-based optimizations. As an example, orthog-onal and redundant basis sets are described in terms of the best approximation based on the SSIM. For the SSIM-Mean, a special case is calculated. Otero, et al. in [11], replace the standard norm-based fidelity measures with SSIM in the optimization problems for image restoration. In their work, the signals (images) are considered to be zero mean. The authors introduce a framework for utilizing the SSIM as a fidelity measure in a wide range of imaging applications. Additionally, they discuss a number of novel numerical strategies that can be employed to solve some standard and original imaging tasks based on the framework. In [35], the authors propose a multi-exposure image fusion algorithm that optimizes a new objective quality measure, known as the color MEF structural sim-ilarity (MEF-SSIM$_c$). Rather than predefining a systematic computation structure for MEF, they search for the image that optimizes MEF-SSIM$_c$ directly in the space of all images.

The general procedure, Various techniques, and categorizations of image fusion are re-viewed below.

3.1   Image Fusion Techniques

In image fusion, the objective is to attain a more representative and detailed output image. An image fusion procedure generally involves several steps that assist in achieving this goal [17]. Several image fusion techniques will be reviewed in the following.

3.1.1   Spatial Based Techniques

Image fusion (IF) can be classified into three groups: spatial, frequency domain, and deep learning methods. In the spatial technique, the pixel values of the input images are manipulated in order to achieve the desired result [7]. Image fusion techniques based on discrete transforms are considerably more prevalent than pyramid-based fusion tech-
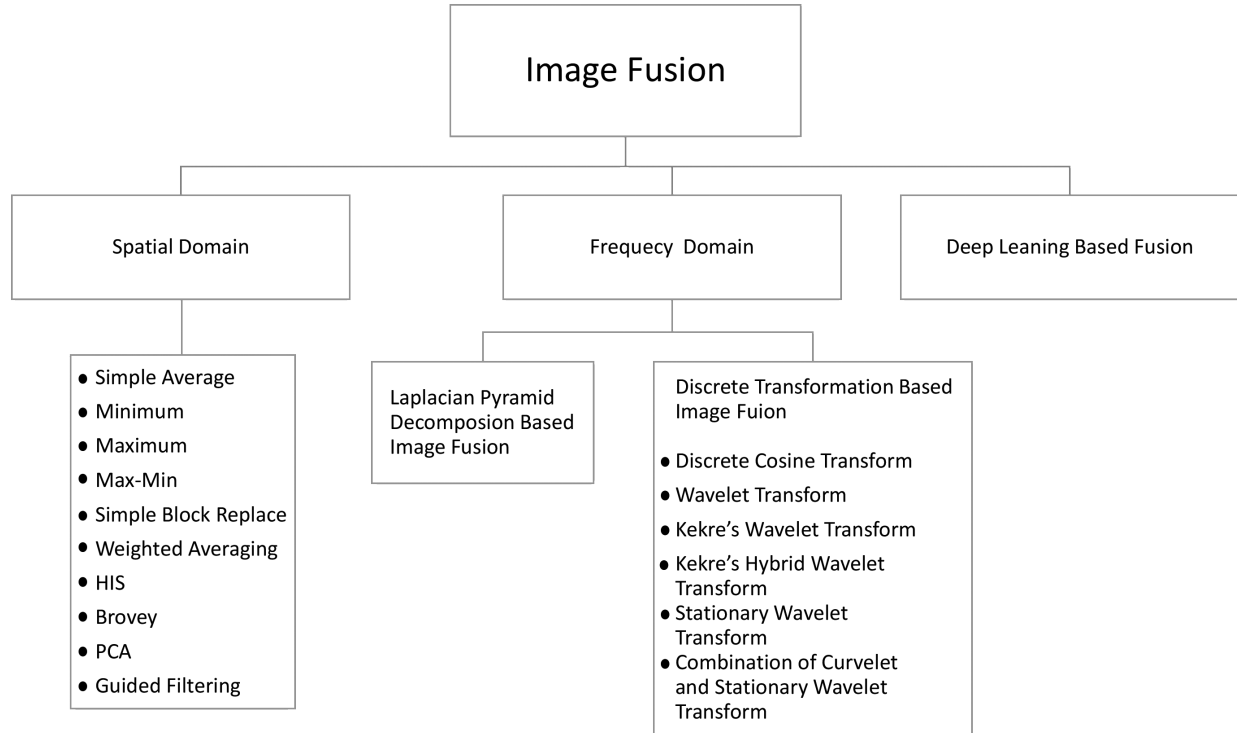
Figure 3.1: Image Fusion Techniques [34].

niques. Figure 3.1 illustrates the different types of IF techniques [54].

There are several spatial based fusion methods, including Max-Min, Minimum, Maximum, Simple Average and Simple Block Replace [32, 33].

3.1.1.1   Simple Average

In this method, images are combined by averaging the pixels. The technique applies to all regions of the image, and it works well if the images are taken from the same type of sensor. Good results can be achieved if the images have high brightness and contrast [34].

3.1.1.2   Minimum Technique

This method selects the pixels with the lowest intensity value and produces a fused image [32]. It is used for darker images [3].

3.1.1.3   Maximum Technique

A fused image is created by selecting the pixels with high intensity from images [7].

3.1.1.4   Max–Min Technique

The resultant combined image was generated by averaging the smallest and largest pixels from the entire source images.

3.1.1.5   Simple Block Replace Technique

All images of pixel values are added together and the block average is calculated. The algorithm is based on pixel neighboring block images.

3.1.1.6   Weighted Averaging Technique

The weights were assigned to every pixel in the source images. A weighted sum of each pixel value in the source images produces the resultant image [69]. By using this method, the output image will be more reliable in terms of detection.

3.1.1.7   Hue Intensity Saturation (HIS)

Based on this technique, the Red-Green-Blue image is converted into HIS components, and then intensity levels are divided using a panchromatic image. Intensity information is contained in spatial and hue and saturation information is contained in spectral. It performs in the bands and has three multispectral bands Red–Green–Blue (RGB) of low resolution. As a final step, the HIS space is converted to RGB space in order to yield the fused image [7]. Combining the features of the images is a very straightforward process, and the result is an image with very high spatial quality. It is the most effective method in remote sensing images, but its major disadvantage is that it only involves three bands [57].

3.1.1.8 Principal Component Analysis (PCA)

It is a statistical method based on orthogonal transformation to convert a set of observations of a possibly correlated variable into principal components which are linearly uncorrelated variables. Spectral degradation and color distortion are two of the principal drawbacks of PCA [2].

3.1.1.9 Guided Filtering

The operator works similarly to the admired bilateral filter in the sense that it smoothes and preserves boundaries. The performance near the boundaries has been enhanced. It is related to a Laplacian matrix in a hypothetical way. It is a fast and non-estimated linear time algorithm, whose density is independent of the mask size. In graphics and computer vision applications, this filter is more efficient and effective due to its ability to join upsampling, eliminate haze, smooth details and reduce noise [27].

3.1.2 Frequency Domain

By applying these techniques, multiscale coefficients from input images can be decomposed [84]. The frequency method can be used to handle spatial distortion.

3.1.2.1 Laplacian Pyramid Fusion Technique

For multi-resolution image fusion, it uses the interpolation sequence and Gaussian pyramid. An improved IF technique was presented by Saleem et al. in which a contrast pyramid transform was employed on multi-source images [12]. The disadvantage of this method is that it does not have sufficient extraction ability, but this can be overcome by decomposing multi-scales. Further, Li et.al. developed a gradient pyramid multi-source IF method that achieves a high band coefficient through the use of gradient direction operators [2].

3.1.2.2   Discrete Transform Fusion Method

Composite images are taken discrete transforms based fusion. As a first step, colored images are separated into RGB (Red-Green-Blue) components. A discrete transformation on the images is applied, and then the average of multiple images is computed [32].

3.1.2.3   Discrete Cosine Transform (DCT)

There are various types of DCT in image fusion, such as DCTma (DCT magnitude), DCTcm (DCT contrast measure), DCTch (DCT contrast highest), DCTe (DCT energy), and DCTav (DCT average) [74]. There is no improvement with this technique when the block size is less than 8×8. A straightforward and basic method of fusing images in the DCT domain is DCTav. DCTe and DCTma methods perform well in image fusion. It is a straightforward technique that can be used in applications that require factual time [34].

3.1.2.4   Discrete Wavelet Transform (DWT) Method

As a result of the DWT method, the two or more images are decompdsed into several high and lo wfrequency bands [8]. By producing a high signal-to-noise ratio with lower spatial resolution, this method minimally distorted the spectral distortion in the resultant fused images as compared to the pixel-based method. A fusion method based on wavelets performed better than a fusion method based on spatial domains for minimizing color distortions [33].

3.1.2.5   Kekre's Wavelet Transform (KWT) Method

The Kekre Wavelet Transform method derives from Kekre's transforms [38]. This method can be used for multiple images and the fused image is superior to other methods [34].

3.1.2.6   Kekre's Hybrid Wavelet Transform (KHWT) Method

A KHWT method has been developed based on hybrid wavelet transforms. A number of authors have suggested that the Kekre-Hadamard wavelet method provides more brightness. A hybrid kekre-DCT wavelet method produces good results in terms of results [34].

3.1.2.7   Stationary Wavelet Transform (SWT) Method

Stationary Wavelet Transform overcomes the disadvantage of translation invariance of the DWT method [37]. The SWT method is derived from the DWT method. The method is a new type of wavelet transform that is translation invariant. Enhanced analysis of image facts is provided by this method [34].

3.1.2.8   Curvelet Transform Method

In terms of time and frequency, SWT has a better characteristic.This can lead to a good result when devising in a smooth manner. Second-generation Curvelet is a new multi-scale transform that overcomes the problems associated with wavelet methods when representing directions of boundaries in the images [6, 15, 53, 62].

3.1.3   Deep Learning

Deep Learning is another technique that is widely used for image fusion in many domains. Several deep learning-based image fusion methods have been presented, showing a variety of advantages, including multi-focus image fusion, multi-exposure image fusion, multi-modal image fusion, Multi-Spectral (MS) image fusion, and Hypers-Pectral (HS) image fusion. There are several recent advances related to deep learning-based images discussed in [46]. To enhance the results of segmentation, deep learning techniques, and case-based reasoning are used with image fusion. Due to the ability of deep learning-based models to extract the most important features automatically from data without human intervention, deep learning methods are becoming increasingly popular for image fusion.

Fusion Techniques	Advantages	Disadvantages
CNN	Able to extract features and representation can learn most elective features from training data without any human intervention	High computational cost
CSR	Compute sparse representation of an entire image shift-invariant representation approach elective in details preservation less sensitive to mis-registration	Need a lot of training data
SAE	Two phase-based training mechanisms have a high potential when the scale of labeled data for supervised learning is limited	If you don't have a good GPU they are quite slow to train (for complex tasks)

Table 3.1: Shows deep learning-based image fusion methods [34]

The models are also able to characterize a variety of complex relationships between input and targeting data. The three deep learning models most commonly used in image fusion are the Convolutional Neural Network (CNN), the Convolutional Sparse Representation (CSR), and the Stacked Autoencoder (SAE). Table 3.1 summarizes the advantages and disadvantages of various deep learning-based image fusion algorithms.

3.1.4 Image Fusion Categorization

It refers to a process that combines the source image and the reference image into one image. In order to achieve the required goal of the fusion objective, different authors anticipated diverse techniques. In order to achieve the required fusion objective, different authors anticipated a variety of techniques. We discuss major classes of such methods below, which include a single sensor, multi-sensor, multimodal, multiview, multi-focus, and multi-temporal [34].

3.1.5 Single Sensor

In order to create an image with the best possible information, a number of images are combined. For example, human operators are not able to perceive desired objects in different lighting conditions and noisy environments, which can be highlighted in the final fused image. This type of system is inadequate as a result of the use of imaging sensors in many sensing areas. The resolution of images is limited by the sensors and the conditions in which the system is operated in its dynamic range. In general, visible band sensors (digital cameras) are suitable for illuminated scenes in daylight but are not suitable for poorly illuminated scenes at night or in fog or rain, which are unfavorable conditions [67].

3.1.6 Multi Sensors

The multi-sensor image fusion is able to overcome the limitations of the one-sensor image fusion since it combines images from multiple sensors in order to produce a compound image. In order to obtain the final image from individual images, a digital camera is accompanied by an infrared (IR) camera. Infrared cameras are suitable for use in inadequately illuminated environments, and digital cameras are suitable for day-light views. It is used in machine vision, medical imaging, robotics, and object detection. It is mainly used to resolve the combined information of multiple images [67].

3.1.7 Multi-View Fusion

There are diverse views at a similar time in multi-view images. Alternatively, it is called mono-modal fusion [67]. It was not always possible to achieve acceptable performances with existing methods, particularly when one of the estimates is poor quality; In this case, they cannot discard it in this case [45, 51, 66].

3.1.8   Multi-Modal Fusion

It refers to combining images from one or more imaging modalities in order to improve image quality. There are several types of models, such as multispectral, panchromatic, infrared, remote sensing, and visible [34].

3.1.9   Multi-Focus Fusion

This method is effective for combining information from a number of images with similar insight features into one wide-ranging image. A compound image is more informative than a single image [47]. A better visual quality is achieved [34]

3.1.10   Multi-Temporal Fusion

The multitemporal fusion process captures the same scene at different times. For the estimation of the occurrence of changes on the ground, long- and short-term observations are required. For a given area, remote sensing images are obtained at diverse times because of revisit observation satellites. Detecting land surface variations across a wide geographic area requires multitemporal images [34].

Chapter 4

Methodology

In this section, A general formulation of the SSIM Mean and its derivative is presented. The SSIM Mean formula is considered a formula for the fusion of images. To combine images, we use the SSIM Mean formula as a fidelity term in the optimization problem. The framework for solving the problem numerically is presented.

To maximize SSIM, we aim to minimize

$$T(\mathbf{x}, \mathbf{y}) = 1 - \text{SSIM}(\mathbf{x}, \mathbf{y}). \tag{4.1}$$

Note that, in contrast to the SSIM, $T(\mathbf{x}, \mathbf{y})$ may be considered a measure of structural dissimilarity between $\mathbf{x}$ and $\mathbf{y}$. The maximization of $\text{SSIM}(\mathbf{x}, \mathbf{y})$ corresponds to the minimization of $T(\mathbf{x}, \mathbf{y})$ in 4.1 ($T(\mathbf{x}, \mathbf{y}) = 0$ when $\mathbf{x} = \mathbf{y}$).

4.1 Introduction

As described in [28] a measurement image $\mathbf{y} = \mathcal{H}(\mathbf{x}) + \mathbf{n}$ is typically inferred using a degradation operator $\mathcal{H}$ and an additive noise term $\mathbf{n}$, where $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$, and $\mathcal{H} \in \mathbb{R}^{n \times m}$. Here it is assumed that $\mathbf{x}$ and $\mathbf{y}$ are vectorized versions, by row stacking, of original 2-D original and distorted images x and y, respectively.

It is generally assumed that the operator $\mathcal{H}$ is known or can be approximated. By

solving the following minimization problem, an approximation of the ideal image x can be obtained. Given a series of images y_i, the problem in general is

$$\arg\min_{\mathbf{x}} \quad \sum_{i=1}^{n} \mathbf{p}_i \; \mathrm{T}(\mathcal{H}_i\mathbf{x}, \mathbf{y}_i) + \lambda g(\mathbf{x}), \qquad (4.2)$$

where $\mathbf{p_i}$ is associated weights, $\mathcal{H} \in \mathbb{R}^{\mathbf{n}\times\mathbf{m}}$ is the degradation operator, $\mathbf{x} \in \mathbb{R}^{\mathbf{m}}$ and $\mathbf{y} \in \mathbb{R}^{\mathbf{n}}$ are vectorized versions of images x and y, $g : \mathbb{R}^{\mathbf{n}} \to \mathbb{R}$ is a regularization functional, which is often defined to be convex, and $\lambda > 0$ is its corresponding regularization parameter. Notice in the above inverse problem as motivated by [28], data fidelity is represented by $\mathrm{SSIM}(\mathcal{H}\mathbf{x}, \mathbf{y})$ instead of the typical sum of squared differences of $\mathcal{H}\mathbf{x}$ and $\mathbf{y}$. Generally, when λ is strictly greater than zero, the regularization term is employed to force the optimal solution to reside within a predefined "target" space, allowing for a *priori* information regarding the solution to be used. Regularization can be particularly useful in the scenario where $\mathcal{H}^T\mathcal{H}$ is weakly conditioned or rank-deficient, in which case it can contribute to well-defined and stably computable solutions. In any case, because the first term in 4.2 is not convex, the whole cost function is not either convex or concave. Therefore, there is no guarantee that a unique global minimizer exists for 4.2. However, it remains possible to develop efficient numerical methods capable of convergent to either a locally or globally optimal solution, as it is demonstrated in the [60].

4.2 SSIM Mean

The problem that we are interested to solve consists of approximating a weighted set of points either on a line or in multidimensional space with a single number or vector. When it comes to statistical inference, the weights represent probabilities and the result is referred to as a point estimator. In image processing, points are interpreted as image patches, which are aggregated into an average image by the estimator. Perceptually relevant averages are desirable, this would mean that the averaged image represents best the content of the set of patches visually. As a result, SSIM, which estimates maximum

structural similarity estimators, is of interest here.

4.2.1 Derivative of SSIM

In this section, the SSIM Mean in terms of alternative expression of SSIM is discussed in chapter2. To solve the fusion image formulated by the SSIM using a gradient-based method, the derivative of the new expression which is based on the mean of vectors and the vector norm function, needs to be computed. A comprehensive explanation is provided on how to formulate SSIM Mean using the expression presented in Chapter 2 and how to calculate its derivatives, as outlined in Appendix 6.2.

The following problem is solved. Given $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ a set of n vectors which their size is $\mathbf{n} \times 1$ and $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ a set of weights, $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_n\}$ a set of degradation operators which its size is $\mathbf{n} \times \mathbf{m}$, find $\mathbf{x}$ which size is $\mathbf{m} \times 1$ such that maximizing

$$\sum_{i=1}^{n} \mathbf{p}_i \ \text{SSIM}(\mathcal{H}_i \mathbf{x}, \mathbf{y}_i). \tag{4.3}$$

By substituting the new expression of SSIM (Equation 2.29), the equation 4.3 can be represented as follows,

$$\sum_{i=1}^{n} \mathbf{p}_i \ \left(\frac{2 \ \overline{\mathcal{H}_i \mathbf{x}} \ \overline{\mathbf{y}_i} + \varepsilon_1}{\overline{\mathcal{H}_i \mathbf{x}}^2 + \overline{\mathbf{y}_i}^2 + \varepsilon_1}\right)$$
$$\left(\frac{\left\|\mathcal{H}_i \mathbf{x} - \mathbf{1}^T \overline{\mathcal{H}_i \mathbf{x}} + \mathbf{y}_i - \mathbf{1}^T \overline{\mathbf{y}_i}\right\|_2^2 - \left\|\mathcal{H}_i \mathbf{x} - \mathbf{1}^T \overline{\mathcal{H}_i \mathbf{x}}\right\|_2^2 - \left\|\mathbf{y}_i - \mathbf{1}^T \overline{\mathbf{y}_i}\right\|_2^2 + \varepsilon_2}{\left\|\mathcal{H}_i \mathbf{x} - \mathbf{1}^T \overline{\mathcal{H}_i \mathbf{x}}\right\|_2^2 + \left\|\mathbf{y}_i - \mathbf{1}^T \overline{\mathbf{y}_i}\right\|_2^2 + \varepsilon_2}\right). \tag{4.4}$$

Note that $\overline{\mathcal{H}_i \mathbf{x}}$ and $\overline{\mathbf{y}}$ are scalar terms. They are considered as column vectors, and set $\overline{\mathcal{H}_i \mathbf{x}} = \mathbf{1}^T \overline{\mathcal{H}_i \mathbf{x}}$ and $\overline{\mathbf{y}} = \mathbf{1}^T \overline{\mathbf{y}}$ where $\mathbf{1} = \begin{bmatrix} 1 \cdots 1 \end{bmatrix}$ is a $1 \times \mathbf{m}$ vector. The cost function from which we take the derivative is

$$P = \sum_{i=1}^{n} \mathbf{p}_i S_1 S_2, \tag{4.5}$$

where,

$$S_1 = \frac{2 \ \overline{\mathcal{H}_i \mathbf{x}} \ \overline{\mathbf{y}_i} + \varepsilon_1}{\overline{\mathcal{H}_i \mathbf{x}}^2 + \overline{\mathbf{y}_i}^2 + \varepsilon_1}, \tag{4.6}$$

$$S_2 = \frac{\left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}} + \mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i}\right\|_2^2 - \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 - \left\|\mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i}\right\|_2^2 + \varepsilon_2}{\left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \left\|\mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i}\right\|_2^2 + \varepsilon_2}. \tag{4.7}$$

By using the quotient rule, we have

$$\frac{\partial p}{\partial \mathbf{x}} = \sum_{i=1}^{n} \mathbf{p}_i \left[\frac{\partial S_1}{\partial \mathbf{x}}S_2 + \frac{\partial S_2}{\partial \mathbf{x}}S_1\right], \tag{4.8}$$

where

$$\frac{\partial S_1}{\partial \mathbf{x}} = \frac{\frac{2}{n}\mathcal{H}_i{}^T\mathbf{1}^T(\overline{\mathbf{y}_i} - \overline{\mathcal{H}_i\mathbf{x}}S_1)}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)}, \tag{4.9}$$

$$\frac{\partial S_2}{\partial \mathbf{x}} = \frac{2\mathcal{H}_i{}^T(\mathbf{I} - \frac{1}{n}\mathbf{J_{n\times n}})\left[(\mathbf{y}_i - \mathbf{1}^T\ \overline{\mathbf{y}_i}) - (\mathbf{I} - \frac{1}{n}\mathbf{J_{n\times n}})\mathcal{H}_i\mathbf{x}S_2\right]}{(\left\|\mathbf{y}_i - \mathbf{1}^T\ \overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\ \overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2)}, \tag{4.10}$$

such that,

$$\mathbf{1}^T\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 \cdots 1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \mathbf{J_{n\times n}}.$$

and $\mathbf{I}$ is the identity matrix whose size is $\mathbf{n} \times \mathbf{n}$.

4.3 Gradient-Descent Method for SSIM-based Optimization

This section provides the solution to the optimization the problem in 4.2. An optimal solution to the optimization problem in equation 4.2 is the image that compromises between the distorted images and the one obtained using the numerical methods in the maximal SSIM sense. The solution of equation 4.2 is achieved by utilizing one of the optimization algorithms.

A gradient-based method is used here to solve this problem. In unrestricted optimization problems, Gradient Descent (GD) optimization algorithms are frequently used as black box optimizers. A gradient-based algorithm uses the gradient's objective function information to determine the minimizer/maximizer cost function with each iteration. GD

is an iterative optimization algorithm of the first order. According to the theory, the steepest descent direction should be taken with repeated steps along the opposite side of the gradient at the current stage (or an approximate gradient) of the function. In contrast, stepping in the direction of the gradient will result in a local maximum of that feature. In this case, the method is known as gradient descent [26]. During the gradient descent algorithm, the magnitude of parameter updates is scaled based on the learning rate. The following algorithm for solving the problem 4.2 is proposed. Furthermore, scipy.optimize.minimize package is used to solve the problem. Further discussion on this topic will be provided below.

It is important to note that the problem under consideration is solved using numerical methods. Consequently, the selection of the starting point and parameters presents limitations in the proposed method. Empirical observations suggest that the arithmetic mean is a suitable choice for the starting point. Additionally, starting from one of the input images can be considered. However, it is crucial to acknowledge the possibility of getting trapped in local minima due to the non-convex nature of the SSIM function. The parameter of the gradient-based algorithm is chosen manually.

Algorithm 1 Gradient Descent Method For SSIM-based Optimization

---

   Initialize $\mathbf{x} = \mathbf{x}_0$;

   take α: *Learning rate ($\alpha > 0$); L:number of the iteration*

   for 1 to L **do**

      Compute Gradient of the cost function $Q(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{p}_i \ \mathrm{T}(\mathcal{H}_i\mathbf{x}, \mathbf{y}_i) + \lambda g(\mathbf{x})$,;

      repeat $\mathbf{x} = \mathbf{x} - \alpha \ Q'(\mathbf{x})$

      until stopping criterion is satisfied;

   end for

   return x

Algorithm 2 Line Search Method Method For SSIM-based Optimization

---

   Initialize $\mathbf{x} = \mathbf{x}_0$;

   take t: *Learning rate ($t > 0$) ; L: the number of iteration; $\alpha \in (0, 0.5)$; $\beta \in (0, 1)$*

   for 1 to L **do**

      while $Q(\mathbf{x}_0 - t \ Q'(\mathbf{x})) > Q(\mathbf{x}_0) + \alpha t Q^T(\mathbf{x})(-Q'(\mathbf{x}))$

      $t = \beta t$

      repeat $\mathbf{x} = \mathbf{x} - t Q'$

      until stopping criterion is satisfied;

   end for

   return x

4.4 Checking Derivatives

In optimization, an erroneous derivative is one of the many traps. An easy way to check the implementation of a derivative is to use the following test. To this end, let f be a multivariate function $f : R^n \longrightarrow R$ and let $v \in R^n$ be an arbitrary vector in the Taylor expansion

$$f(\mathbf{x} + h\mathbf{v}) = f(\mathbf{x}) + hdf(\mathbf{x})v + O(h^2). \tag{4.11}$$

The Matrix $\mathbf{A}$ is the derivative of f if and only if the difference between $||f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x}) - h\mathbf{A}\mathbf{v}||$ is quadratic in h. That difference is calculated by the function check derivative (see, e.g., [55]). The result of applying the check-Derivative function to the SSIM derivative is shown in 4.1. In this case, the result is consistent since the difference is quadratic with respect to h.
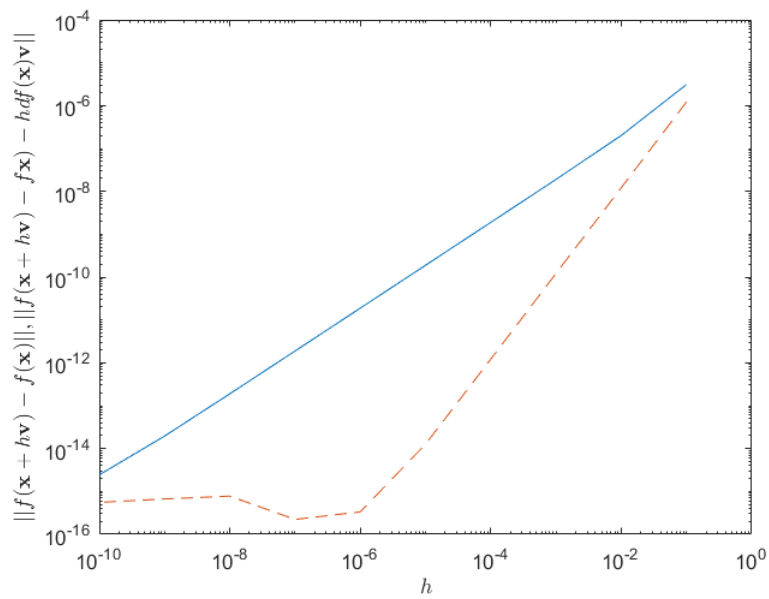
Figure 4.1: CheckDerivative result for implementation of SSIM derivatives: linear decay (solid line) for $||f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})||$ and quadratic decay (dashed line) for $||f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x}) - hdf(\mathbf{x})\mathbf{v}||$ up to machine precision is shown on a logarithmic scale.

Chapter 5

Experiments and Results

In this Chapter, we report and discuss the experimental results of the proposed method. To compare our results with Brunet's, we present a few examples where we've found SSIM-Mean using both Brunt's and our approaches. In the following section, we will discuss the design of the experiment for the fusion of images. Afterward, numerical and visual results will be presented in order to evaluate the performance of the proposed methods.

5.1   One Dimensional Case

In [5], Brunet addresses the problem of maximizing SSIM Mean for a set of points and presented a solution. To evaluate the proposed method, an example mentioned by Brunet on page 97 of his thesis, considering one-dimensional data, is replicated. Furthermore, an additional one-dimensional example is presented.

Brunet discusses the problem of finding the maximum SSIM Mean between two given vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ in his thesis [5], i.e. the problem is finding $\mathbf{x}$ such that

$$\arg\max_{\mathbf{x}} \quad \text{SSIM}(\mathbf{x}, \mathbf{y}_1) + \text{SSIM}(\mathbf{x}, \mathbf{y}_2). \tag{5.1}$$

The objective is to determine the maximum value of the following function, which is obtained by substituting the given values and parameters into the formula of SSIM.

$$f(\mathbf{x}) = \frac{2\ \mathbf{x}\ (\mathbf{y}_1) + \varepsilon_1}{\mathbf{x}^2 + \mathbf{y}_1{}^2 + \varepsilon_1} + \frac{2\ \mathbf{x}\ (\mathbf{y}_2) + \varepsilon_1}{\mathbf{x}^2 + \mathbf{y}_2{}^2 + \varepsilon_1}. \tag{5.2}$$

Note that the mean of a scalar is itself (if $\mathbf{x}$ is scalar then $\bar{\mathbf{x}} = \mathbf{x}$).

As previously discussed, the objective of maximizing SSIM between two given vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ as used by Brunet is equivalent to finding the minimum value of $T(\mathbf{x}, \mathbf{y}) = 1 - \mathrm{SSIM}(\mathbf{x}, \mathbf{y})$ which is used in the proposed calculations.

The problem involves finding the minimum value of SSIM Mean between two given vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ by solving equation 4.2. The regularization term in equation 4.2 is disregarded as the focus here is on finding the local maximum of SSIM Mean. To accomplish this objective, it is assumed that $\mathbf{p}_i = 1$ and $\mathcal{H}_i = \mathbf{I}$, where $\mathbf{I}$ represents the identity matrix. Hence,

$$\arg\min_{\mathbf{x}}\ \ T(\mathbf{x}, \mathbf{y}_1) + T(\mathbf{x}, \mathbf{y}_2). \tag{5.3}$$

The problem is solved numerically using scipy.optimize, which offers various optimization methods. The L-BFGS-B method in Scipy-optimization is used when $\epsilon_1 = 0$ in this case. Figures 5.1 and 5.2 provide a visual presentation of the function $\mathbf{x}$ for two cases, i) for $\mathbf{y}_1 = 0.1$ and $\mathbf{y}_2 = 0.4$ and ii) for $\mathbf{y}_1 = 0.05$ and $\mathbf{y}_2 = 0.8$ when $\epsilon_1 = 0$. In the first case, a local maximum of 0.2 is achieved, which is consistent with Brunet's result. In the second case, a maximum value of 0.69 is obtained when starting from 0.8, and a maximum value of 0.057 is obtained when starting from 0.05. These results are in accordance with the findings reported in Brunet's thesis. Figures 5.1 and 5.2 are identical to those presented in Brunet's thesis. Note that the Matlab code of Brunet's work is provided in his thesis
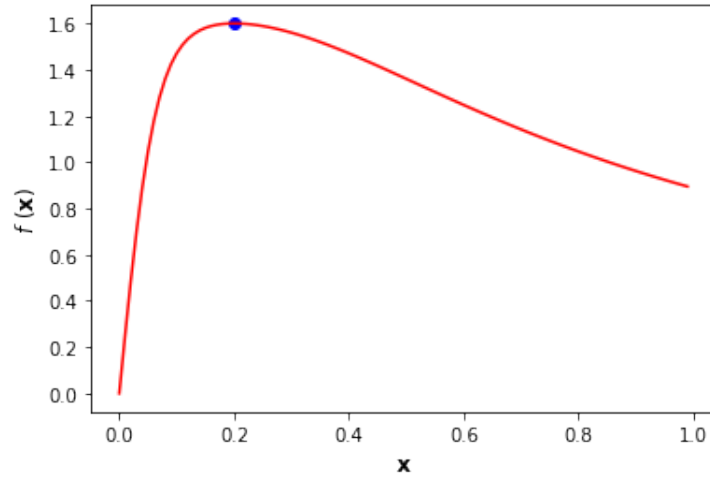
Figure 5.1: Virtual presentation of function $f(\mathbf{x}) = \text{SSIM}(\mathbf{x}, 0.1) + \text{SSIM}(\mathbf{x}, 0.4)$. The blue point represents the local maximum obtained by the L-BFGS-B method in the Scipy-optimization package when the start point is 0.1.
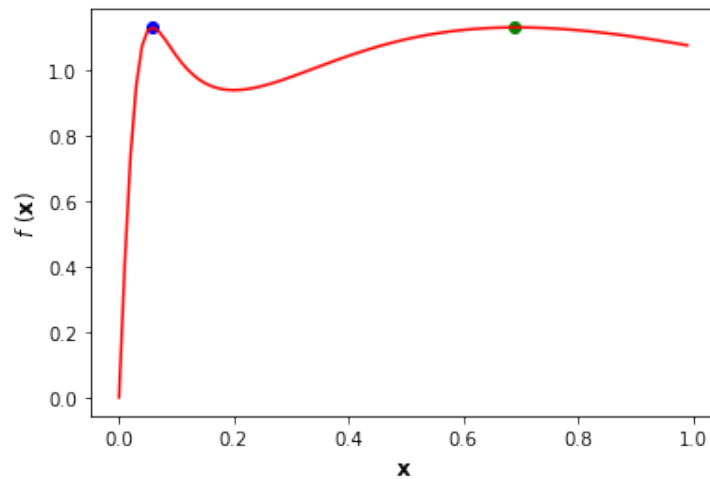
Figure 5.2: Virsual presentation of function $f(\mathbf{x}) = \text{SSIM}(\mathbf{x}, 0.05) + \text{SSIM}(\mathbf{x}, 0.8)$. The blue and green points represent the local maximum obtained by the L-BFGS-B method in the Scipy-optimization package when the starting point is 0.05 and 0.8 respectively.

In the second experiment, the local maximum of SSIM Mean is found for four points, $\mathbf{y}_1 = 0.07$, $\mathbf{y}_2 = 0.6$, $\mathbf{y}_3 = 28$, $\mathbf{y}_4 = 65$, i.e. the local maximum of $h(\mathbf{x})$ is determined
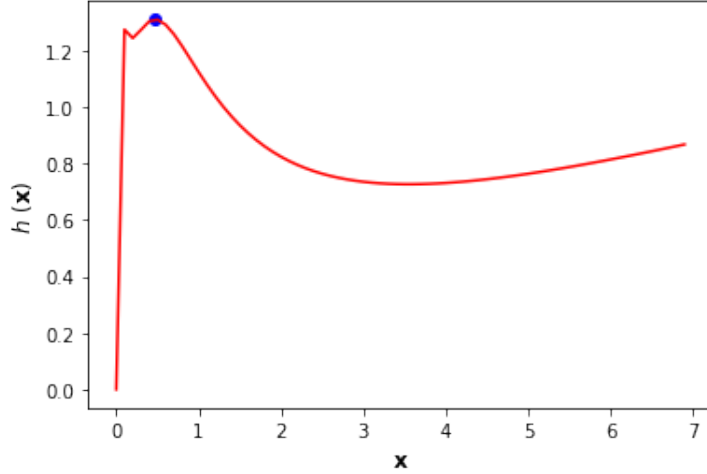
Figure 5.3: Virtual presentation of function $h(\mathbf{x}) = \frac{2\mathbf{x}(0.07)}{\mathbf{x}^2+0.07^2} + \frac{2\mathbf{x}(0.6)}{\mathbf{x}^2+0.6^2} + \frac{2\mathbf{x}(28)}{\mathbf{x}^2+28^2} + \frac{2\mathbf{x}(65)}{\mathbf{x}^2+65^2}$. The blue point represents the local maximum obtained by the L-BFGS-B method in the Scipy-optimization package when the start point is 0.6.

where,

$$h(\mathbf{x}) = \frac{2\mathbf{x}(0.07)}{\mathbf{x}^2 + 0.07^2} + \frac{2\mathbf{x}(0.6)}{\mathbf{x}^2 + 0.6^2} + \frac{2\mathbf{x}(28)}{\mathbf{x}^2 + 28^2} + \frac{2\mathbf{x}(65)}{\mathbf{x}^2 + 65^2}. \tag{5.4}$$

Note that ϵ_1 is set to zero, similar to the previous example.

Both the proposed method and Brunet's method can be used to solve the problem. By using the Matlab code provided in Brunet's thesis, a local maximum of 0.47 for $h(\mathbf{x})$ is obtained. Similarly, by using the L-BFGS-B method in the Scipy-optimization package with a starting point of 0.6, the same result is achieved. Figure 5.3 illustrates the graphical representation of the function $h(\mathbf{x})$.

In the subsequent sections, the experimental design will be presented, followed by the application of images to the experiments. To demonstrate the flexibility of the proposed method, examples will be provided under two different assumptions: (1) assuming $\mathcal{H} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and (2) assuming $\mathcal{H} = \mathcal{D}_z$, which represents a local averaging followed by a downsampling factor of z transformation. It is important to note that any degradation operator can be applied to the proposed method. In Equation 4.2, total variation, which is commonly used as a regularization term in image processing, is

applied as a regularization term. However, it is possible to substitute other regularization functions instead. Moreover, if necessary, additional terms can be incorporated into Equation 4.2.

When images are used as inputs in the problem, local SSIM (MSSIM) instead of SSIM is applied in Equation 4.2. Consequently, the algorithms 1 and 2 require the use of the derivative of MSSIM instead of SSIM. To address this, the derivative of MSSIM as presented in [59] is used for solving the problem.

5.2   SSIM-Total Variation Regularization

Naturally, different sets of constraints and regularization terms yield different SSIM-based imaging problems. As anticipated, by choosing different types of regularization terms and degradation operators, a wide variety of SSIM-based imaging tasks can be performed. Total Variation(TV) is selected as regularization in this section for the whole of the experiments. TV regularization tends to preserve edges in general, and in certain cases does indeed preserve edges exactly. It should be noted that maximizing the similarity between two images (vectors) is equivalent to minimizing dissimilarity between them. Let $y_1, y_2, \ldots, y_n$ be a set of n degraded images which their size is $\mathtt{m} \times \mathtt{p}$ and $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ a set of associated weights across patches, $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_n\}$ a set of degradation operators which its size is $\mathtt{n} \times \mathtt{m}$, find x which size is $\mathtt{n} \times \mathtt{p}$ such that

$$\arg\min_x \sum_{i=1}^{n} \mathbf{p}_i \, \mathrm{TM}(\mathcal{H}_i x, y_i) + \lambda \|\nabla x\|_{TV}, \tag{5.5}$$

where $\|\nabla x\|_{TV} = \sum_{\mathsf{x}} \sqrt{\|\nabla x(\mathsf{x})\|_2^2}$ is isotropic total variation regularization image x and $\mathrm{TM}(\mathcal{H}_i x, y_i) = 1 - \mathrm{MSSIM}(\mathcal{H}_i x, y_i)$. To use the gradient-based algorithm described in 1 and 2, we must calculate the derivative of TV regularization as

$$Grad\left(\|\nabla x\|_{TV}\right) = -\nabla \cdot \left(\frac{\nabla x}{\|x\|}\right) = -\frac{x_{\mathsf{xx}} x_{\mathsf{y}}^2 - 2x_{\mathsf{x}} x_{\mathsf{y}} x_{\mathsf{xy}} + x_{\mathsf{yy}} x_x^2}{(x_{\mathsf{x}}^2 + x_{\mathsf{y}}^2)^{3/2}}. \tag{5.6}$$

Considering the finite difference is taken by assuming Neumann boundary conditions. Then, evolve the PDE (with a small nonzero factor added to the fraction term's denominator, a which is called "fudge parameter" in this case, to prevent the term from becoming singular)

$$Grad\left(\|\nabla x\|_{TV}\right) = -\nabla \cdot \left(\frac{\nabla x}{\|x\|}\right) = -\frac{x_{\mathsf{xx}}x_{\mathsf{y}}^2 - 2x_{\mathsf{x}}x_{\mathsf{y}}x_{\mathsf{xy}} + x_{\mathsf{yy}}x_x^2}{a + (x_{\mathsf{x}}^2 + x_{\mathsf{y}}^2)^{3/2}}. \tag{5.7}$$

It should be noted that in the proposed method, the scaling of $\mathbf{p}_i$ can be adjusted based on the confidence level of each measurement. Additionally, $\mathcal{H}_i$ can be a different degradation operator for each measurement. However, for the sake of simplicity and to evaluate the method's performance, $\mathcal{H}_i$ is considered $\mathcal{H}$ and $\mathbf{p}_i$ is considered 1 in this case. In future studies, different values for $\mathbf{p}_i$ and alternative degradation operators can be considered. Additionally, all parameters are manually chosen in a range that is suitable for the experiment.

5.3   Fusion of Degraded Images In The Case that $\mathcal{H} = \mathbf{I}$

5.3.1   Denoising Experiment

5.3.1.1   Cameraman Test Image

This experiment aims to perform denoising by applying a set of noisy images and obtaining a denoised image that combines these noisy images.

The optimization problem 5.5 is solved using the proposed gradient-descent algorithm 2. In this case, the denoising task is applied to four images of the test image "Cameraman" corrupted with additive white Gaussian noise (AWGN). Additional information regarding various types of noise can be found in the standard textbook "Digital Image Processing" [22].

By adding AWGN with a mean of zero and variance of 0.005 four times to the Cameraman image, four noisy images of it are obtained, where each image corresponds to one of

y_1, y_2, y_3, and y_4 in equation 5.5. In Figure 5.7, the first row displays one of our inputs, which corresponds to the noisy images.

Since equation 5.5 is numerically solved using algorithm 2, it requires a starting point to initiate the algorithm. In the following, the results obtained from two different starting points are compared and discussed.

1. **The starting point is the arithmetic mean of noisy Cameraman images.**

   The starting point in this case is the arithmetic mean of these four images (the third image in the first row in Figure 5.7). It corresponds to x in the equation 5.5 and is used in the first step of the algorithm. The values chosen for the parameters of the line search algorithm are 0.4 and 0.01, while the learning rate is set to 0.01. Additionally, the fudge factor in the derivative of the regularization is assigned a value of 0.001. Table 5.1 shows the PSNR values of the noisy images, which are approximately 23, and their corresponding SSIM values, which are around 0.44..

   When $\lambda = 0$, only the SSIM-Mean term applies in equation 5.5, resulting in an image that closely resembles the arithmetic mean image, which is the starting point. The optimal answer is obtained by gradually increasing the value of λ until the ideal balance between the fidelity term and the regularization term is found. It should be noted that selecting excessively large values for λ leads to a disregard for the fidelity term in equation 5.5. However, in gradient-based methods, there are other factors that influence and control the results, such as the learning rate, the number of iterations, and the fudge parameter.

   The results presented in both Table 5.1 and Figure 5.7 demonstrate that, as expected, using SSIM leads to obtaining a denoised image that visually is superior. In Table 5.1, it is observed that the PSNR value of the arithmetic image used as the starting point is 29.30, with an SSIM value of 0.68. The best result in terms of PSNR and MSE corresponds to the $\lambda = 3$ which its PSNR and MSE are 31.93 and

41.68 respectively. The graph of MSE and PSNR with respect to λ in the selected range is observed in Figures 5.4 and 5.5. In terms of the SSIM, the best result in the chosen range of λ is obtained around $\lambda = 3.5$ based on Figure 5.6. The SSIM value which is associated with $\lambda = 3.5$ is 3.5. As expected, the optimal image is obtained, characterized by a lower MSE value, a higher PSNR value, and a higher SSIM value compared to the individual input images and the starting point. As shown in Figure 5.7, the denoised image is obtained by the combination of a set of noisy images that is superior visually to each of our noisy images and starting point individually.

	MSE	PSNR	SSIM
Noisy image 1	297.94	23.38	0.44
Noisy image 2	303.69	23.30	0.44
Noisy image 3	302.92	23.31	0.44
Noisy image 4	305.88	23.27	0.43
Arithmetic of noisy images	76.37	29.30	0.68
Denoising image-$\lambda = 0$	76.37	29.30	0.68
Denoising image-$\lambda = 2.5$	42.13	31.88	0.83
Denoising image-$\lambda = 3$	**41.68**	**31.93**	0.84
Denoising image-$\lambda = 3.5$	43.33	31.76	**0.85**

Table 5.1: Numerical results for denoising the fusion of noisy Cameraman images. (The starting point is the arithmetic mean of noisy Cameraman images).

Figure 5.4: Change of MSE value with respect to λ. (The starting point is the arithmetic mean of noisy Cameraman images).

Figure 5.5: Change of PSNR value with respect to λ. (The starting point is the arithmetic mean of noisy Cameraman images).

Figure 5.6: Change of SSIM value with respect to λ. (The starting point is the arithmetic mean of noisy Cameraman images).

Figure 5.7: Visual results for the denoising of the test image Cameraman. (a) Original image. (b) One of the noisy images. (c) Arithmetic mean. (d) Denoised image ($\lambda = 0$). (e) Denoised image ($\lambda = 0.1$). (f) Denoised image ($\lambda = 1$). (g) Denoised image ($\lambda = 2.5$). (h) Denoised image ($\lambda = 3$). (i) Denoised image ($\lambda = 3.5$) (The starting point is the arithmetic mean of noisy Cameraman images).

	MSE	PSNR	SSIM
Noisy image 1	305.04	23.28	0.43
Noisy image 2	300.29	23.35	0.44
Noisy image 3	303.72	23.30	0.43
Noisy image 4	302.06	23.32	0.44
Denoising image-$\lambda = 0$	304.98	23.28	0.43
Denoising image-$\lambda = 6$	**94.55**	**28.37**	0.77
Denoising image-$\lambda = 10$	102.16	28.03	**0.82**

Table 5.2: Numerical results for denoising a set of noisy images. (The starting point is one of the noisy Cameraman images).

2. **The starting point is one of the noisy images**

   Here, one of the noisy images is selected as the starting point, and the algorithm 1 is applied. A learning rate of 0.01 and a fudge parameter of 0.001 are used. The values of MSE and PSNR of the starting point are approximately 300 and 23, respectively. As the previous one, when $\lambda = 0$, only the SSIM-Mean term is considered in the equation 5.5. In order to achieve a balance between fidelity and regulation terms, λ is increased. The results show that the optimal results in terms of MSE and PSNR is achieved when $\lambda = 6$, whereas the superior result in terms of SSIM is obtained when $\lambda = 10$. These findings are presented in Table 5.2 and illustrated in the plots of Figure 5.8, 5.9, and 5.10. The visual representation of the denoised image can be observed in Figure 5.11.

   The obtained results align with the initial expectation, which aimed to achieve a denoised image with higher PSNR and SSIM values and a lower MSE by combining multiple noisy images. Also, it is seen that the results are biased toward the starting point which is selected for the algorithm 2.
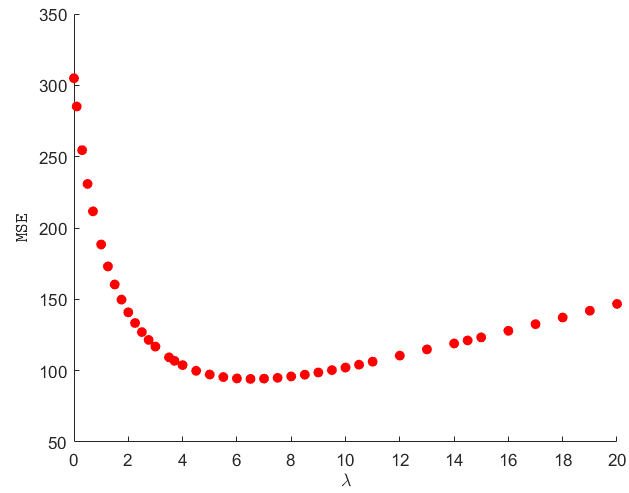
Figure 5.8: Change of MSE value with respect to λ. (The starting point is one of the noisy Cameraman images).
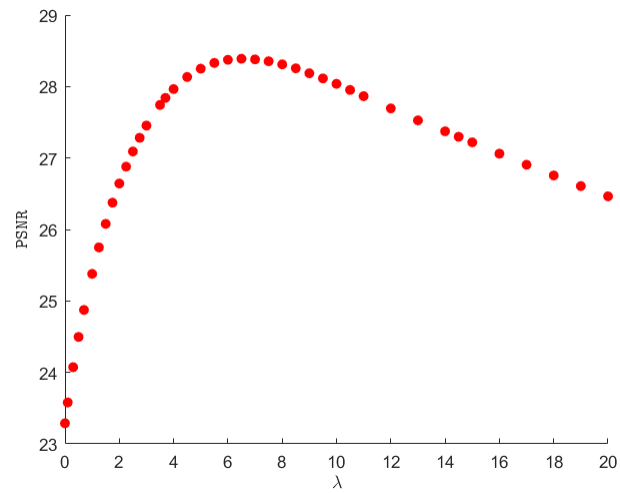
Figure 5.9: Change of PSNR value with respect to λ.(The starting point is one of the noisy Cameraman images).

Figure 5.10: Change of SSIM value with respect to λ. (The starting point is one of the noisy Cameraman images).

(a)   (b)   (c)

(d)   (e)   (f)
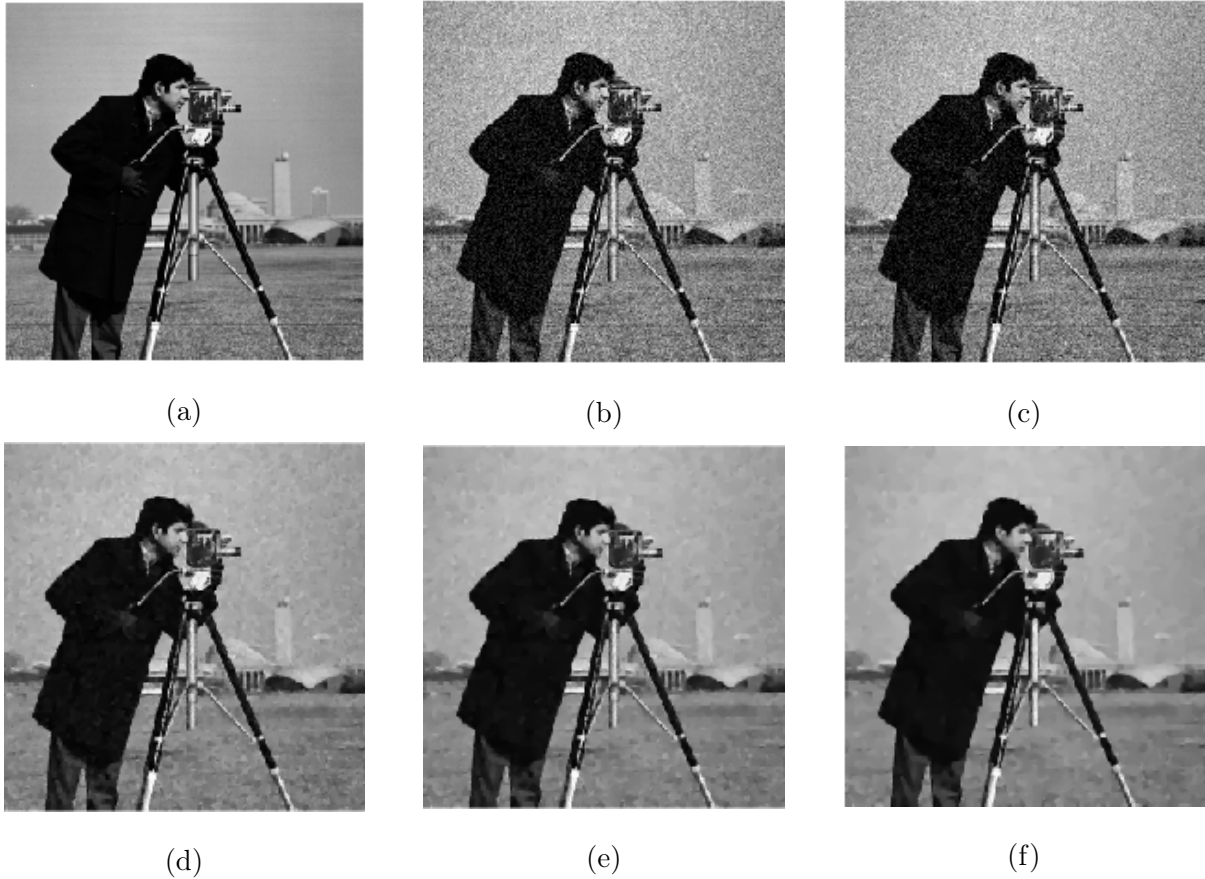
Figure 5.11: Visual results for the denoising of the test image Cameraman. (a) Original image. (b) Noisy image 1. (c) Noisy image 3 . (d) Denoised image ($\lambda = 6$). (e) Denoised ($\lambda = 8$). (f) Denoised image ($\lambda = 10$). (The starting point is one of the noisy Cameraman images).

5.3.1.2   Mandrill Test Image

Illustrations of the results for the "Mandrill test image" are included in all sections of the experiments, demonstrating that the proposed method can be applied to any image.. In this experiment, four Gaussian noisy images are created in the same manner as for the Cameraman test image. Denoised images are obtained by fusing noisy images using equation 5.5. The equation is solved using algorithm 1. The learning rate and fudge parameter of regularization here, are 0.01 and 0.001 respectively. The starting point is

the arithmetic mean of four noisy images.

As observed and anticipated, the applicability of the proposed method extends to various images, including the Mandrill test image. The plots of MSE, PSNR, and SSIM as functions of λ (refer to Figures 5.12, 5.13, and 5.14) within the chosen range of λ, along with the corresponding numerical results presented in Table 5.3, demonstrate that the optimal results in terms of MSE, PSNR, and SSIM within the specified range are achieved when $\lambda = 1$. The visual result is shown in Figure 5.15.

Although Mandrill's image has more details, the denoised image which is superior virtually compare to our input noisy images individually and the starting point is obtained.

	MSE	PSNR	SSIM
Noisy image 1	323.15	23.03	0.65
Noisy image 2	324.94	23.01	0.65
Noisy image 3	323.99	23.02	0.65
Noisy image 4	324.58	23.01	0.65
Arithmetic of noisy images	81.11	29.03	0.85
Denoising image-$\lambda = 0.01$	81.05	29.04	0.86
Denoising image-$\lambda = 0.1$	77.36	29.24	0.86
Denoising image-$\lambda = 1$	**67.01**	**29.86**	**0.88**
Denoising image-$\lambda = 2$	80.12	29.09	0.87
Denoising image-$\lambda = 2.5$	89.49	28.61	0.86

Table 5.3: Numerical results for denoising a set of noisy Mandrill images. (The starting point is the arithmetic mean of noisy Mandrill images).
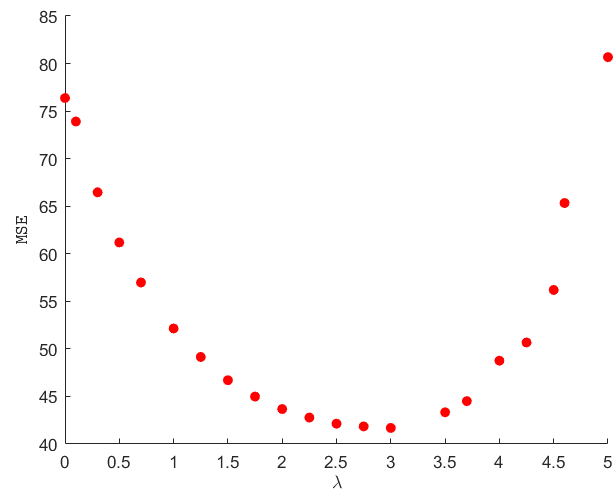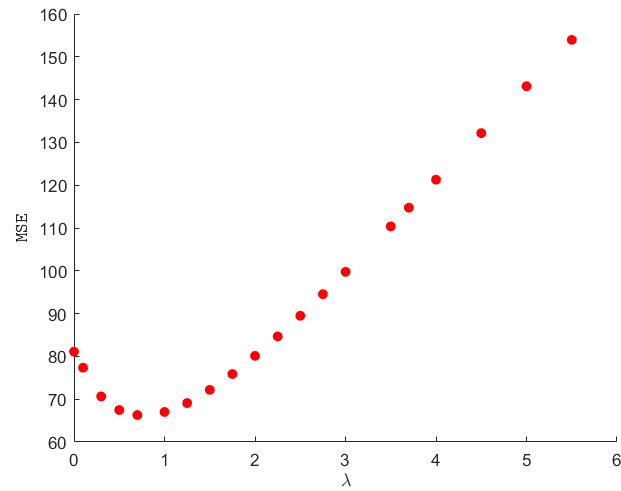
Figure 5.12: Change of MSE value with respect to λ. (The starting point is the arithmetic mean of noisy Mandrill images).
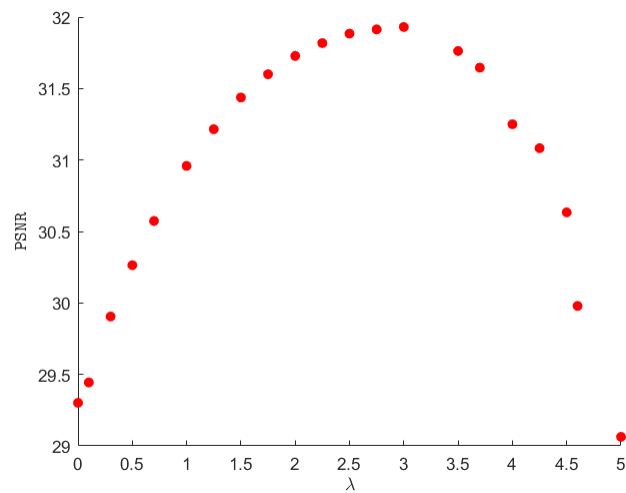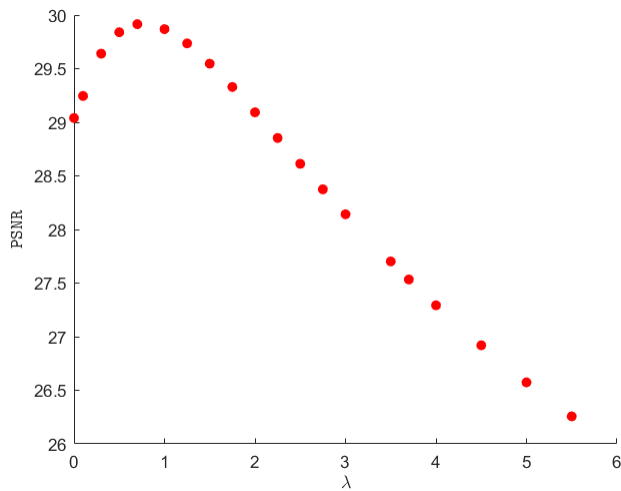

Figure 5.13: Change of PSNR value with respect to λ. (The starting point is the arithmetic mean of noisy Mandrill images).
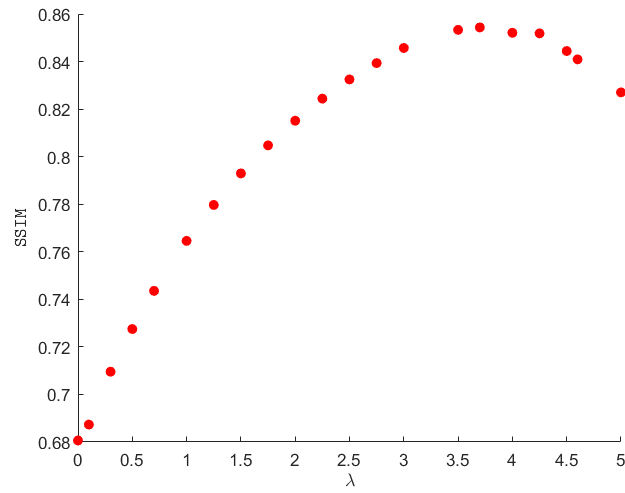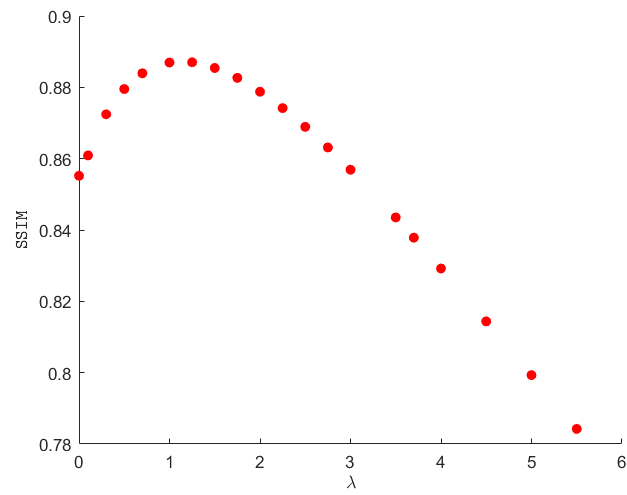
Figure 5.14: Change of SSIM value with respect to λ. (The starting point is the arithmetic mean of noisy Mandrill images).
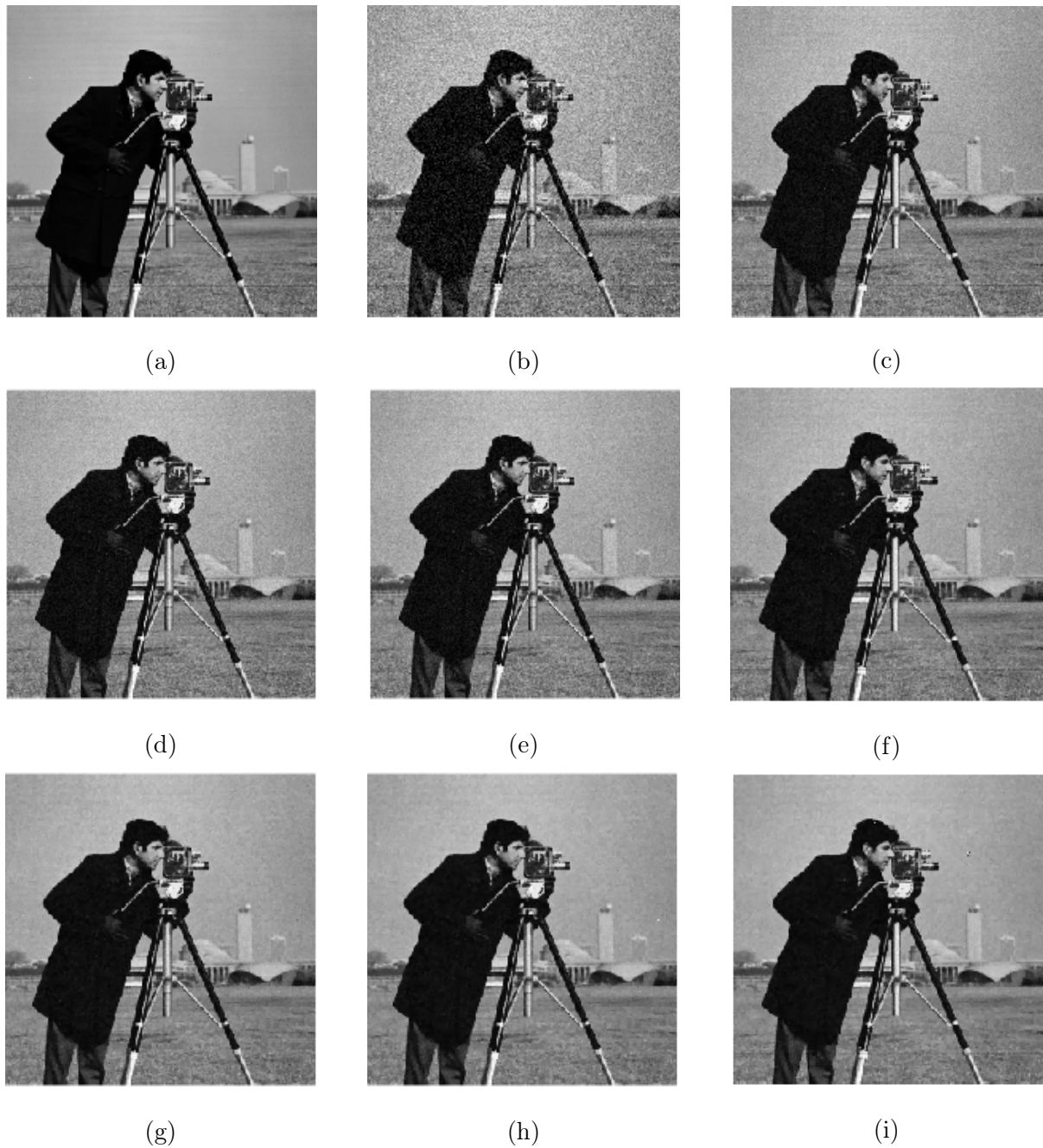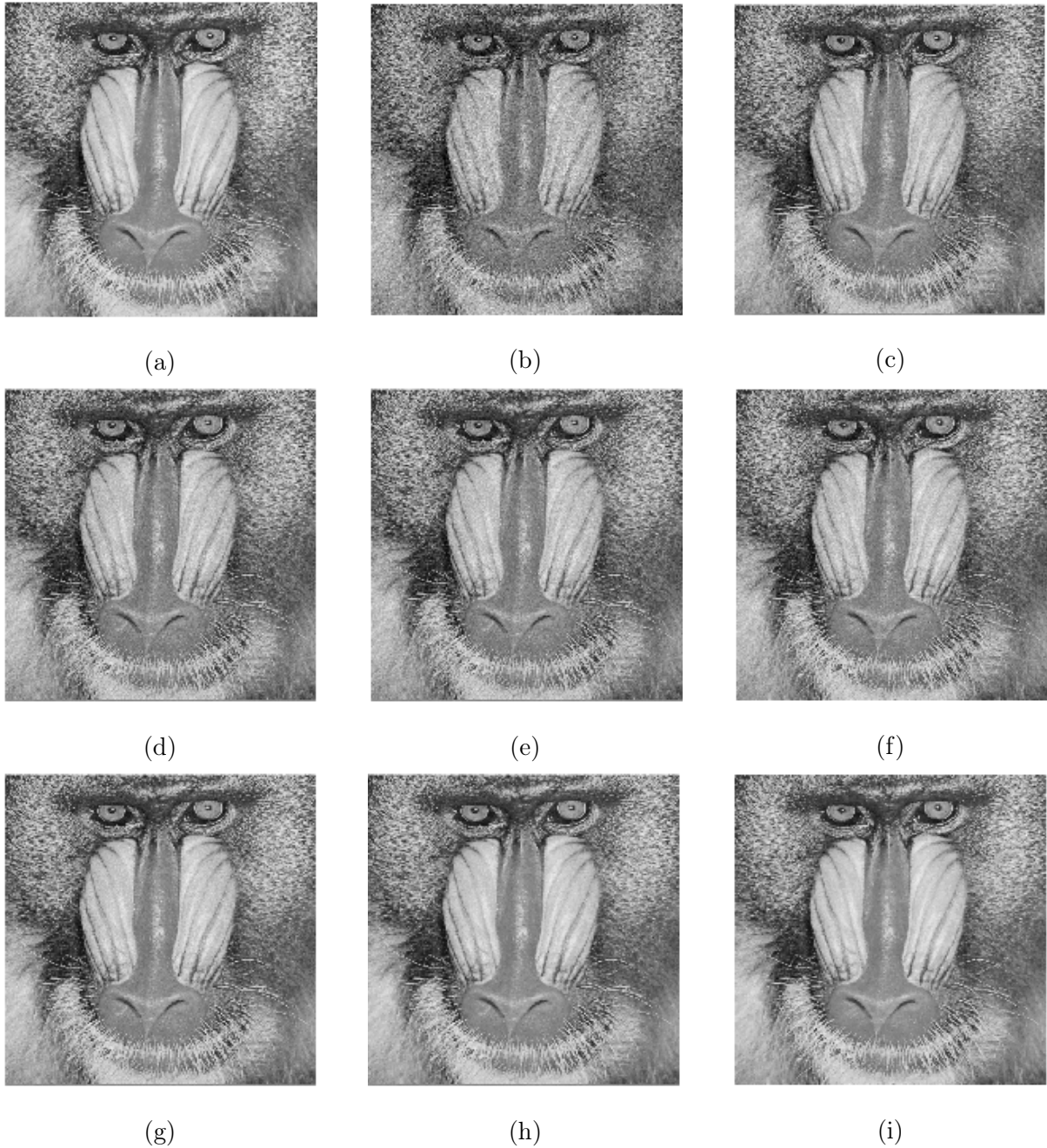
Figure 5.15: Visual results for the denoising of the test image Mandrill. (a) Original image. (b) One of the noisy images. (c) Arithmetic mean. (d) Denoised image ($\lambda = 0$). (e) Denoised image ($\lambda = 0.01$). (f) Denoised image ($\lambda = 0.1$). (g) Denoised image ($\lambda = 1$). (h) Denoised image ($\lambda = 2$). (i) Denoised image ($\lambda = 2.5$). (The starting point is the arithmetic mean of noisy Mandrill images).

5.3.2   Fusion of Different Types of Image Distortion

5.3.2.1   Cameraman Test Image

In order to demonstrate the performance of the proposed method for fusing different types of degradation, in this experiment the image of the Cameraman is degraded as follows

- Mean contrast stretched (Multiplying the original image by 1.2),

- Mean shift (Adding 10 to the original image),

- Blurring (Applying a 10-by-10 filter),

- Salt and pepper noise,

- Gaussian noise.

Note that Gaussian noise with a mean of zero and variance of 0.005 is added to the mean contrast stretched and mean shift images.

Each degraded image corresponds to one of y_1, y_2, y_3, y_4, and y_5 in equation 5.5. The distorted images can be seen in Figure . We use algorithm 1. The learning rate and fudge parameter are considered 0.01 and 0.001 respectively .

Two different starting points are considered and the result is described for each of them.

1. **The starting point is the Gaussian noisy image**

   The Gaussian noisy image is considered as the start point of the algorithm 1. The SSIM term (fidelity term) is the only term in the equation 5.5 when $\lambda = 0$. It is expected that the optimal solution to equation 5.5 would result in a fused image from these five distortion images with the highest value of PSNR and SSIM and the lowest MSE values.

   MSE, PSNR, and SSIM values for our degraded images and some fused images can be seen in Table 5.4. The plots of MSE, PSNR, and SSIM changes as a function

of λ are shown in Figures 5.16, 5.17, and 5.16. According to the table and plots, the superior fused image in the selected range of λ in terms of MSE and PSNR corresponds to $\lambda = 7$ and for SSIM it corresponds to $\lambda = 10$.

The visual result of this experiment can be seen in Figure 5.19. This Figure shows that, as expected, the fused image for $\lambda = 10$ is superior to each of the degraded images.

	MSE	PSNR	SSIM
Gaussian noisy mean contrast stretched	1005.62	18.10	0.44
Gaussian noisy mean-shifted	419.92	21.89	0.42
Blurring	705.04	19.64	0.56
Gaussian noise	301.62	23.33	0.44
Salt and pepper	1027.86	18.01	0.38
Fusion image-$\lambda = 0$	301.50	23.33	0.44
Fusion image-$\lambda = 0.1$	281.10	23.64	0.45
Fusion image-$\lambda = 1$	184.38	25.47	0.53
Fused image-$\lambda = 5$	94.49	28.37	0.75
Fused image-$\lambda = 7$	**92.74**	**28.45**	0.80
Fused image-$\lambda = 10$	101.91	28.04	**0.82**

Table 5.4: Numerical results for the fusion of a set of distortion images. (The starting point is the Gaussian noisy image).

Figure 5.16: Change of MSE value with respect to λ. (The starting point is the Gaussian noisy image).

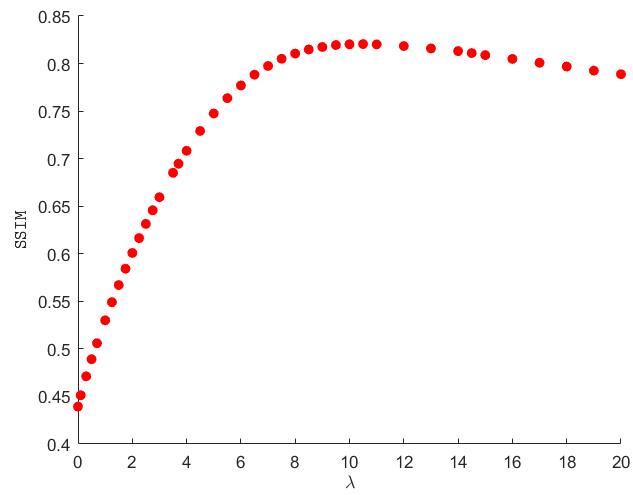Figure 5.17: Change of PSNR value with respect to λ. (The starting point is the Gaussian noisy image).

Figure 5.18: Change of SSIM value with respect to λ. (The starting point is the Gaussian noisy image).

Figure 5.19: Visual results for the fusion of a set of distortion versions of the test image Cameraman. (a) Original image. (b) Gaussian noisy mean contrast Stretched. (c) Gaussian noisy mean shifted. (d) Blurring. (e) Gaussian noise. (f) Salt and pepper noise. (g) Fused image ($\lambda = 5$). (h) Fused image ($\lambda = 7$). (i) Fused image ($\lambda = 10$). (The starting point is the Gaussian noisy image).

2. **The starting point is salt and pepper noisy image**

   In this case, the values of the parameters are the same as in the previous case. The MSE, PSNR and SSIM values of the salt and pepper noisy image are 1027.86, 18.01, and 0.38, respectively. By combing these degraded images, it is aimed to obtain a fused image that is superior to each of them individually.

   It is attempted to make a balance between fidelity and regulation terms by increasing λ in the optimization problem 5.5. The numerical results in Table 5.5 shows that in the selected range of λ the superior fused image obtains when $\lambda = 14$ in terms of MSE, PSNR, and SSIM, i.e. it is clearly seen that the values of PSNR and SSIM are higher than each of our degraded images individually and MSE is lower than each of them individually for $\lambda = 14$ as expected. The visual results are provided in Figure 5.23. The plot of MSE, PSNR and SSIM with respect to λ can be seen in Figures 5.20, 5.21, and5.22, respectively.

	MSE	PSNR	SSIM
Gaussian noisy mean contrast stretched	1005.62	18.10	0.44
Gaussian noisy mean-shifted	419.92	21.89	0.42
Blurring	705.04	19.64	0.56
Gaussian noise	301.62	23.33	0.44
Salt and pepper	1027.86	18.01	0.38
Fusied image-$\lambda = 3.5$	227.06	24.56	0.77
Fused image-$\lambda = 7$	173.54	25.73	0.78
Fused image-$\lambda = 14$	**150.79**	**26.34**	**0.79**

Table 5.5: Numerical results for the fusion of a set of distortion images. (The starting point is the salt and pepper noisy image).
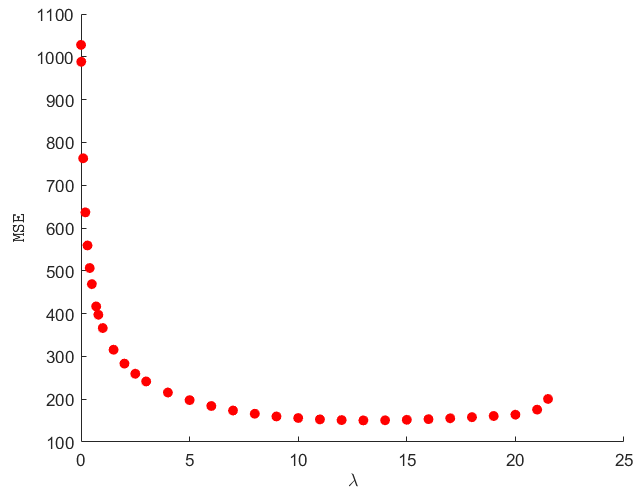
Figure 5.20: Change of MSE value with respect to λ. (The starting point is the salt and pepper noisy image).
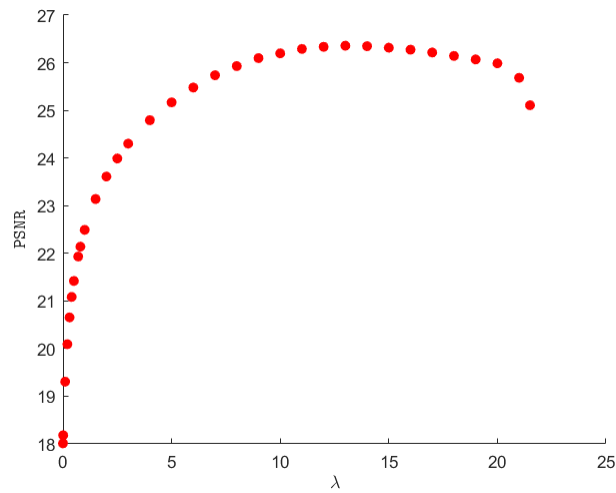
Figure 5.21: Change of PSNR value with respect to λ. (The starting point is the salt and pepper noisy image).
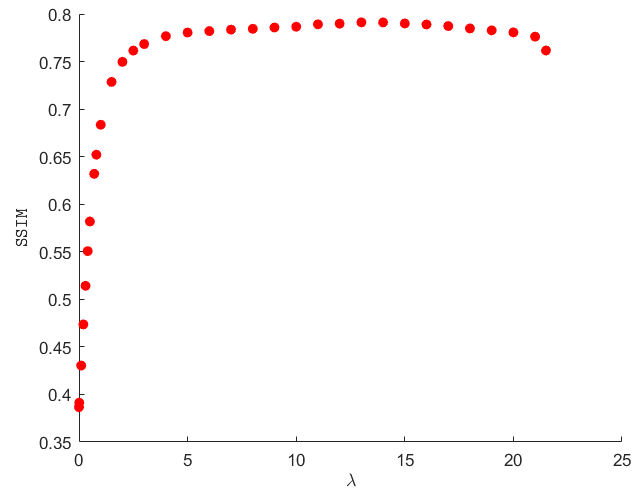
Figure 5.22: Change of SSIM value with respect to λ. (The starting point is the salt and pepper noisy image).

Figure 5.23: Visual results for the fusion of a set of distortion versions of the test image Cameraman. (a) Original image. (b) Gaussian noisy mean contrast Stretched. (c) Gaussian noisy mean shifted. (d) Blurring. (e) Gaussian noise. (f) Salt and pepper noise. (g) Fused image ($\lambda = 3.5$). (h) Fused image ($\lambda = 7$). (i) Fused image ($\lambda = 14$). (The starting point is the salt and pepper noisy image).

5.3.2.2   Mandrill Test Image

As mentioned previously, the Mandrill test image is applied in the same experiment described above, i.e. the Mandrill image is degraded by five different types of distortion, and the goal is to find a fused image that is superior to each of the degradation images individually.

The learning rate and fudge parameter of regularization are respectively 0.01 and 0.001. In this case, the starting point is the Gaussian noisy image. Numerical results, plots of MSE, PSNR, and SSIM against λ, as well as the visual result are presented in the Tables 5.6, and Figures 5.24, 5.25, 5.26 and 5.27, respectively.

	MSE	PSNR	SSIM
Gaussian noisy mean contrast stretched	1045.5	17.93	0.66
Gaussian noisy mean-shifted	423.08	21.86	0.65
Blurring	683.49	19.78	0.300
Gaussian noise	324.31	23.02	0.65
Salt and pepper	910.08	18.54	0.54
Fusion image-$\lambda = 0.01$	323.53	23.03	0.65
Fusion image-$\lambda = 2$	193.44	25.26	0.74
Fused image-$\lambda = 2.5$	186.51	25.42	0.75
Fused image-$\lambda = 3.5$	**182.81**	**25.51**	**0.76**

Table 5.6: Numerical results for the fusion of a set of distortion images. (The starting point is the Gaussian noisy image of Mandril).
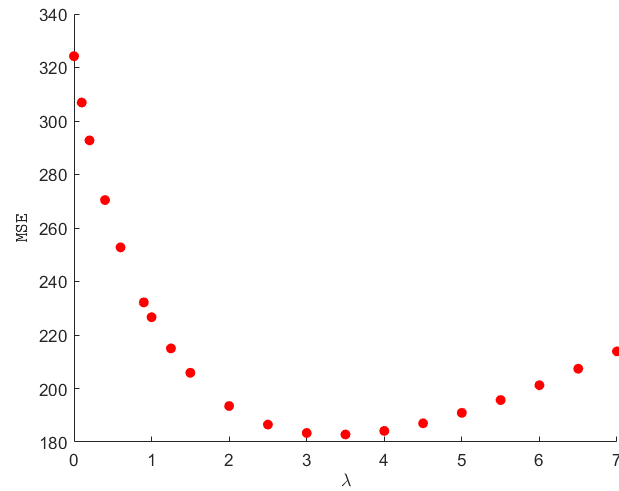
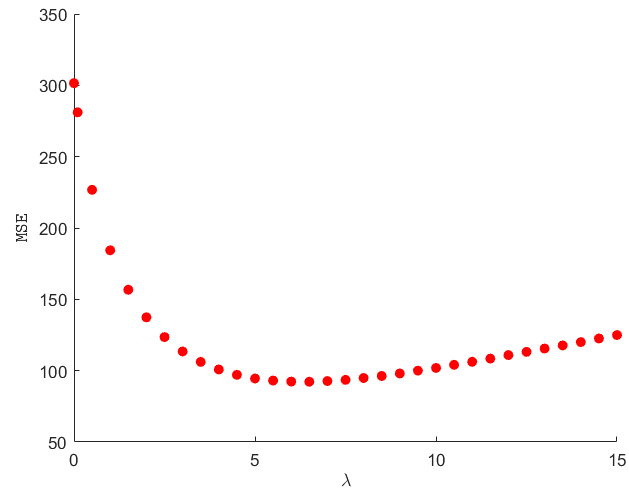Figure 5.24: Change of MSE value with respect to λ. (The starting point is the Gaussian noisy image of Mandril).
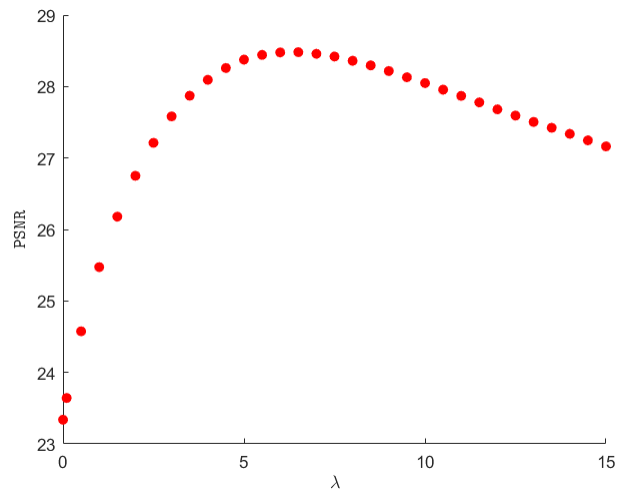

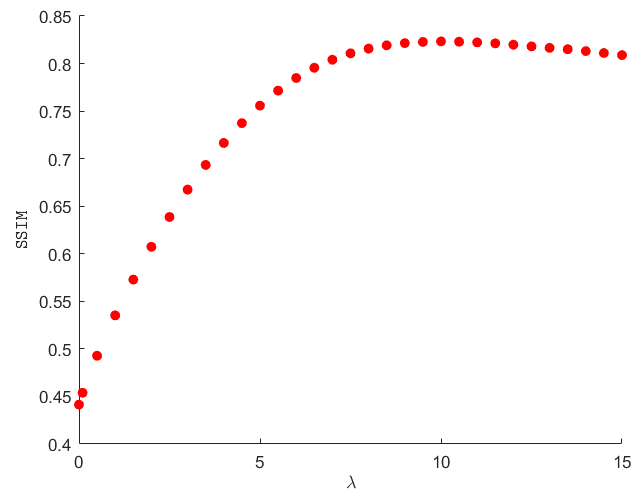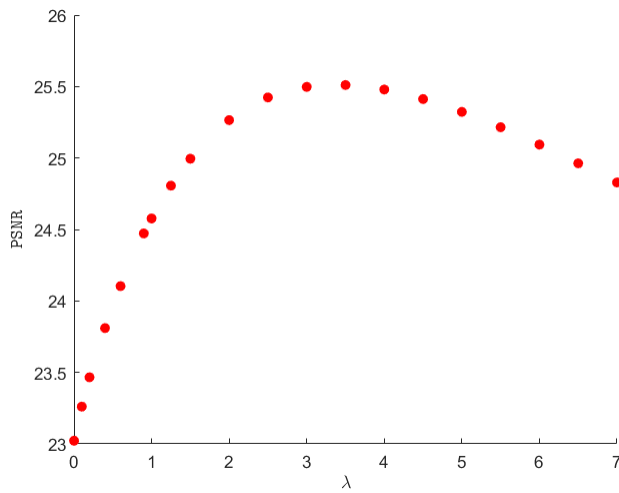Figure 5.25: Change of PSNR value with respect to λ. (The starting point is the Gaussian noisy image of Mandril).
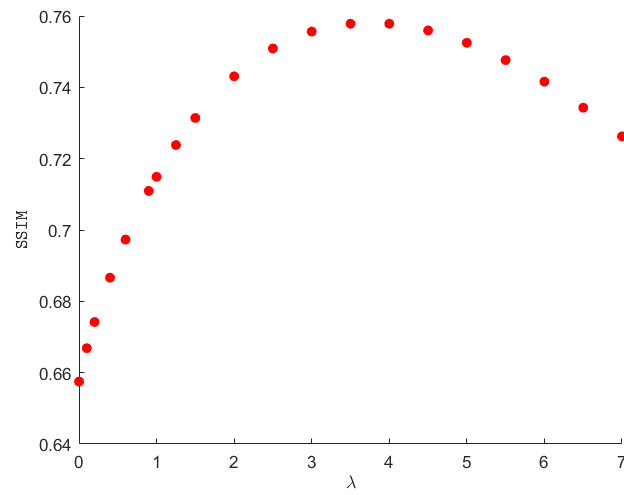
Figure 5.26: Change of SSIM value with respect to λ.(The starting point is the Gaussian noisy image of Mandril).
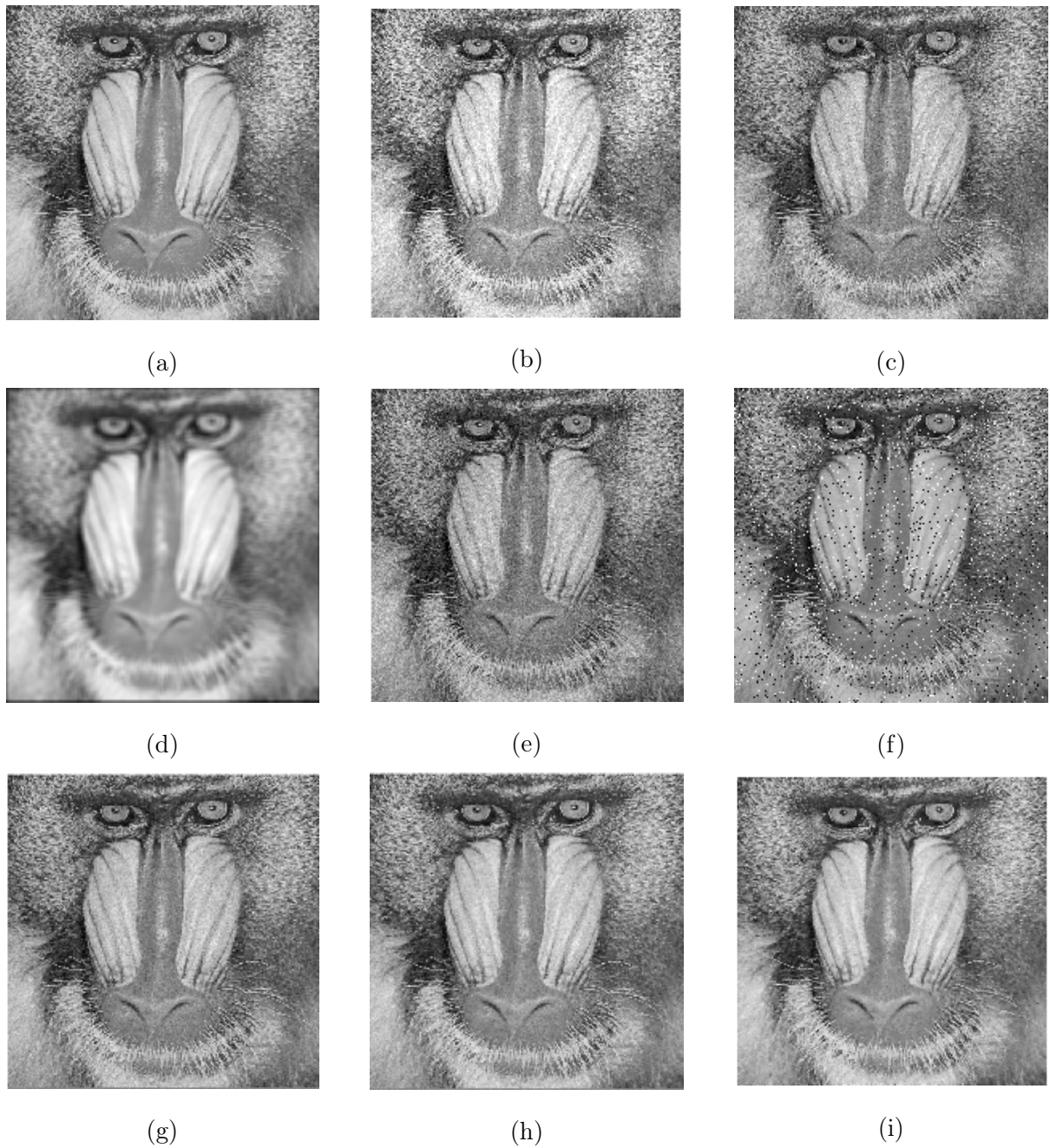
Figure 5.27: Visual results for the fusion of a set of distortion versions of the test image Mandrill. (a) Original image. (b) Gaussian noisy mean contrast Stretched. (c) Gaussian noisy mean shifted. (d) Blurring. (e) Gaussian noise. (f) Salt and pepper noise. (g) Fused image ($\lambda = 2.5$). (h) Fused image ($\lambda = 3.5$). (i) Fused image ($\lambda = 5$). (The starting point is the Gaussian noisy image of Mandril).

5.4 Fusion Degraded Images In The Case That $\mathcal{H} = \mathcal{D}_z$

$\mathcal{D}_z$ in the equation5.5 is local averaging followed by downsampling factor of z . Following are some experiments in which degradation transformation operators other than identity operators are applied. In these cases, the goal is to demonstrate how flexible our methods are with respect to applying any type of degradation operator.

5.4.1 Zooming a Set of Noisy Images

5.4.1.1 Cameraman Test Image

In this experiment, the following conditions and assumptions are considered.

- Test image "Cameraman" is used as a reference image, which has a size of 256×256.

- local averaging followed by a downsampling factor transformation factor of two is taken into account as $\mathcal{H}$ in the equation 5.5.

- distortion happened in the following manner:
  First, local averaging followed by a downsampling transformation factor of two is applied to the original image of the Cameraman, resulting in an image with a size of 128×128Four degraded images are obtained with a size of 128×128 by adding Gaussian noise, whose mean value is zero and variance is 0.005, four times to the downsampled image of Cameraman. These four degraded images are considered as y_1, y_2, y_3, y_4 in the equation 5.5.

- Fudge parameters for the regularization term and the learning rate of the gradient descent algorithm (1) are 0.001 and 0.1.

- In order to start the algorithm we need an initial point. The arithmetic mean of y_1, y_2, y_3, and y_4 is selected as the starting point. Note that, this is an arithmetic mean of resized input images using interpolation. The following two results are

(a) (b)

Figure 5.28: Visual results of one of the noisy images and resized version of it. a) The Cameraman's image is degraded by local averaging followed by a downsampling factor of two and Gaussian noise. (b) The resized version of the distorted image (a) by applying bicubic interpolation.

presented for both bicubic interpolation and nearest interpolation of y_1, y_2, y_3, and y_4.

- Since downsampling occurs in these examples, it is not expected to obtain optimal images whose PSNR and SSIM are significantly higher than the starting point, and each of the distortion images individually. Although there is a small increase in these values. As can be seen from the virtual result, denoising does take place.

1. **Result obtained by using bicubic interpolation to resize the images**

   Figure 5.28 shows one of the noisy images as well as the resized version of it. The arithmetic mean of resized versions of images is used as the starting point of the algorithm 1. The learning rate is 0.01 and the fudge parameter for the regularization is 0.001. With increased λ in 5.5, the denoised images is obtained with higher values of PSNR, SSIM, and lower values of MSE compare to each resized version of y_1, y_2, y_3, and y_4. It can be seen that the lowest MSE and highest PSNR is obtained when

	MSE	PSNR	SSIM
Resize of noisy image 1	348.61	22.70	0.44
Resize of noisy image 2	346.68	22.73	0.44
Resize of noisy image 3	346.09	22.76	0.44
Resized of noisy image 4	351.93	22.66	0.44
Arithmetic mean of 4 Resize of noisy images	197.52	25.17	0.65
Fused image-$\lambda = 0$	197.53	25.17	0.65
Fusion image-$\lambda = 1$	182.59	25.51	0.75
Fusion image-$\lambda = 1.5$	**182**	**25.53**	0.77
Fusion image-$\lambda = 2.5$	186.59	25.42	**0.79**

Table 5.7: Numerical results for the fusion of a set of distortion images. (In the case of bicubic interpolation used to resize the noisy Cameraman images).

$\lambda = 1.5$, and the highest SSIM corresponds to $\lambda = 2.5$ (The numerical result has been compiled in the Table 5.7). The plots of MSE, PSNR, SSIM against λ, and the visual result, are presented in Figures 5.29, 5.30, and 5.31 and 5.32, respectively.
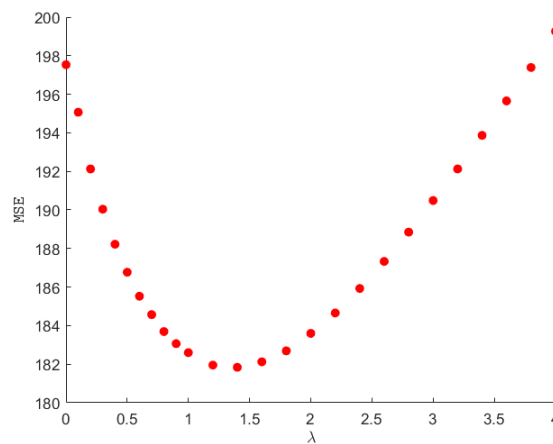
Figure 5.29: Change of MSE value with respect to λ. (In the case of bicubic interpolation used to resize the noisy Cameraman images).

Figure 5.30: Change of PSNR value with respect to λ. (In the case of bicubic interpolation used to resize the noisy Cameraman images).

Figure 5.31: Change of SSIM value with respect to λ. (In the case of bicubic interpolation used to resize the noisy Cameraman images).

Figure 5.32: Visual results for the fusion of a set of distortion of the test image Cameraman. (a) Original image. (b) One of the resized versions of the Gaussian noisy image. (c) The arithmetic mean of four resized versions of the noisy image. (d) Fused image ($\lambda = 0$). (e) Fused image ($\lambda = 1.5$). (f) Fused image ($\lambda = 2.5$) (In the case of bicubic interpolation used to resize the noisy Cameraman images).

2. **Result obtained by using nearest interpolation to resize the images**

   Here, the parameters and starting point are the same as those used in examples that used bicubic interpolation for resizing images. The goal is to combine these distorted images and obtain a fused image that is superior to each distorted image by solving the optimization problem 5.5. This means that it is aimed to obtain a fused image that has both higher PSNR and SSIM and lower MSE as compared to the individual distorted images. The highest PSNR and SSIM values are obtained

when $\lambda = 3.8$, as shown in the plots of 5.33, 5.34, and 5.35 over the selected range of λ. Additionally, the corresponding MSE value is the lowest at $\lambda = 3.8$. Figures 5.8, 5.33, 5.34, 5.35 and 5.36 provide numerical results, plots of MSE, PSNR, SSIM, and visual results, respectively.

	MSE	PSNR	SSIM
Resized version of noisy image 1	490.40	21.22	0.36
Resized version of noisy image 2	488.77	21.23	0.36
Resized version noisy image 3	490.05	21.22	0.37
Resized version of noisy image 4	491.13	21.21	0.36
Arithmetic mean of 4 resized version of noisy images	261.40	23.95	0.58
Fused image-$\lambda = 0$	261.40	23.95	0.58
Fusion image-$\lambda = 0.01$	261.4	23.95	0.58
Fused image-$\lambda = 1$	219.27	24.72	0.69
Fused image-$\lambda = 2.5$	195.07	25.22	0.77
Fused image-$\lambda = 3.8$	**191.48**	**25.30**	**0.79**
Fused image-$\lambda = 4.5$	192.49	25.28	0.78

Table 5.8: Numerical results for the fusion of a set of distortion images. (In the case of nearest interpolation used to resize the noisy Cameraman images).
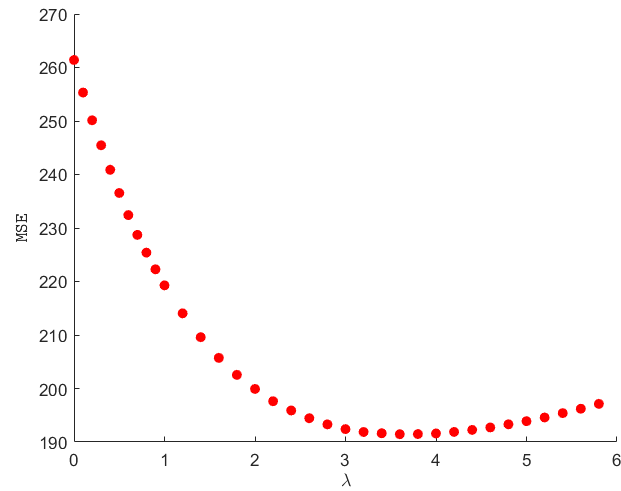
Figure 5.33: Change of MSE value with respect to λ. (In the case of nearest interpolation used to resize the noisy Cameraman images).

Figure 5.34: Change of PSNR value with respect to λ. (In the case of nearest interpolation used to resize the noisy Cameraman images).

Figure 5.35: Change of SSIM value with respect to λ. (In the case of nearest interpolation used to resize the noisy Cameraman images).

(a)   (b)   (c)

(d)   (e)   (f)
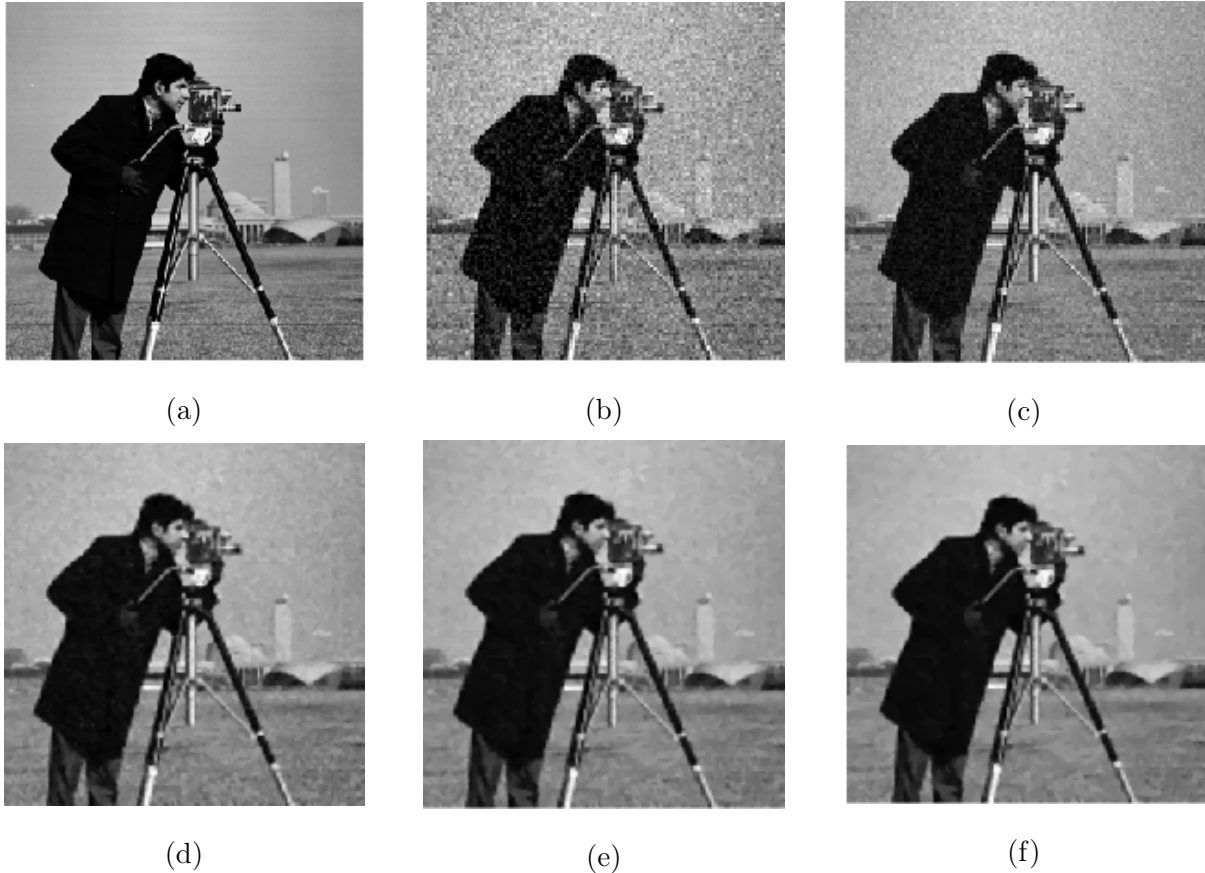
Figure 5.36: Visual results for the fusion of a set of distortion versions of the test image Cameraman. (a) Original image. (b) One of the resize of Gaussian noisy image. (c) The arithmetic mean of four resizes the noisy image. (d) Fused image ($\lambda = 1$). (e) Fusedimage ($\lambda = 2.5$). (f) Fused image ($\lambda = 3.8$) (In the case of nearest interpolation used to resize the noisy Cameraman images).

5.4.1.2   Mandrill Test Image

The previous example is redone by applying the Mandrill test image with the same assumptions, except that in this case, the degradation operator $\mathcal{H}$ in equation 5.5 is set as local averaging followed by a downsampling factor of 4. Figures 5.9, 5.37, 5.38, 5.39 and 5.40 provide numerical results, plots of MSE, PSNR, SSIM with respect to λ in the selected range, and visual results.

	MSE	PSNR	SSIM
Resize of noisy image 1	855.47	18.80	0.27
Resize of noisy image 2	853.90	18.81	0.27
Resize of noisy image 3	856.87	18.80	0.27
Resize of noisy image 4	870.21	18.73	0.27
Arithmetic mean of 4 resize of noisy images	614.78	20.24	0.37
Fusion image-$\lambda = 0$	614.78	20.24	0.37
Fusion image-$\lambda = 1$	599.15	20.35	0.38
Fusion image-$\lambda = 2.5$	583.73	20.46	**0.39**
Fusion image-$\lambda = 4.5$	**577.70**	**20.51**	0.38

Table 5.9: Numerical results for the fusion of a set of distortion images. (In the case of nearest interpolation used to resize the noisy Mandrill images).
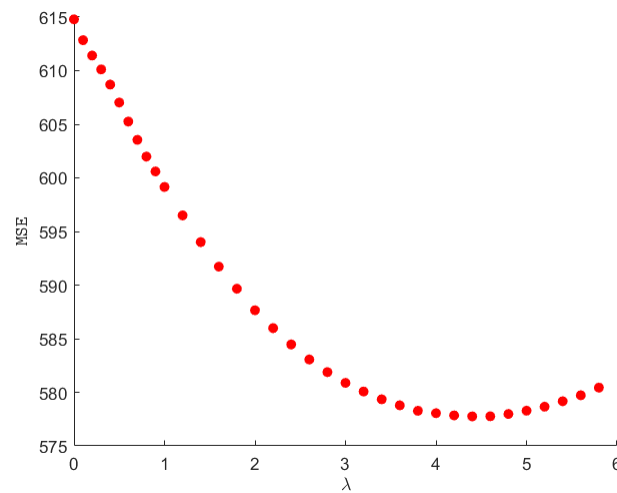
Figure 5.37: Change of MSE value with respect to λ. (In the case of nearest interpolation used to resize the noisy Mandrill images).
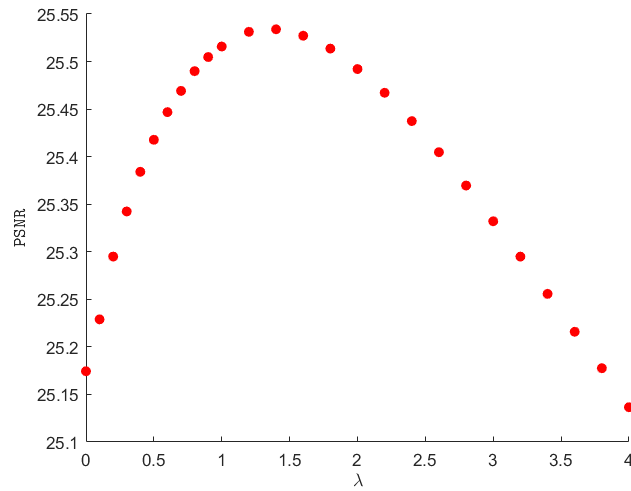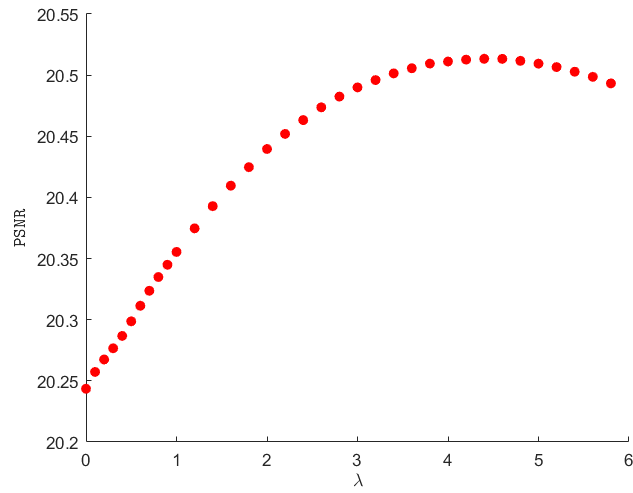
Figure 5.38: Change of PSNR value with respect to λ. (In the case of nearest interpolation used to resize the noisy Mandrill images).
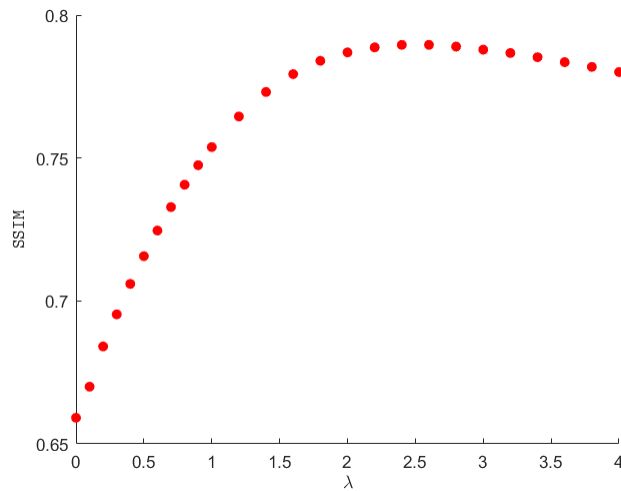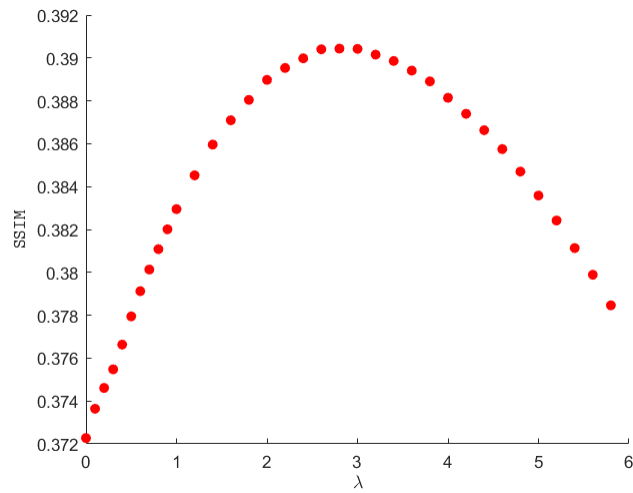

Figure 5.39: Change of SSIM value with respect to λ. (In the case of nearest interpolation used to resize the noisy Mandrill images).
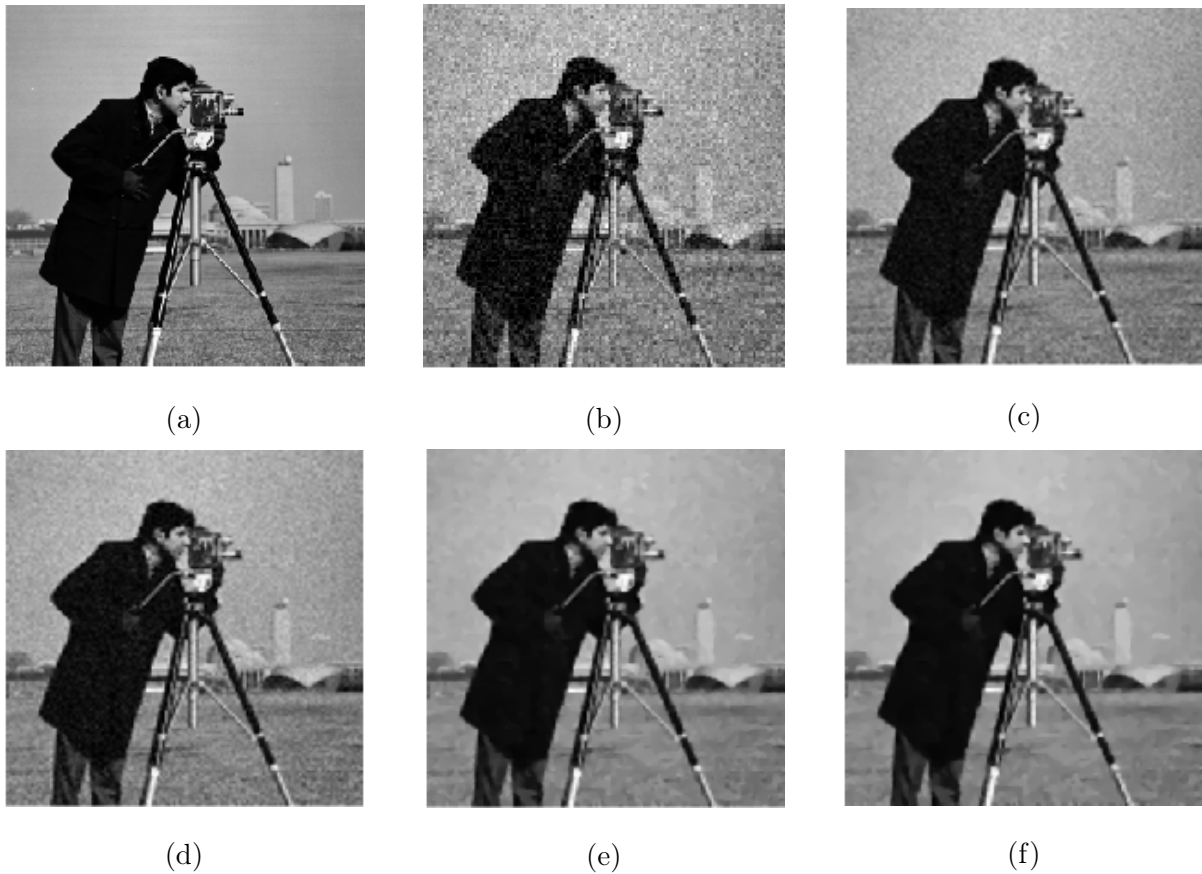
Figure 5.40: Visual results for the fusion of a set of distortion of the test image Mandrill. (a) Original image. (b) One of the resize of Gaussian noisy image. (c) The arithmetic mean of four resizes the noisy image. (d) Fused image ($\lambda = 1$). (e) Fused image ($\lambda = 2.5$). (f) Fused image ($\lambda = 3.8$). (In the case of nearest interpolation used to resize the noisy Mandrill images).

5.4.2   Fusion of Different Types of Image Distortion

5.4.2.1   Cameraman Image

1. **Result obtained by using bicubic interpolation to resize the images**

   In this experiment, the local averaging followed by a downsampling transformation factor of 2 which is associated with $\mathcal{H}$ in the equation 5.5 is applied on the test image "Cameraman" and then it is degraded by five different distortions similar

to experiment 5.3.2 (except for blurring, a 5-by-5 filter is applied). Each degraded image corresponds to one of y_1, y_2, y_3, y_4, and y_5 in equation 5.5. Images are resized using bicubic interpolation. Both degraded images and resized versions of them are shown in Figure 5.41.

All the other values for the parameters are the same as the experiment 3. A resized version of the mean shift image is used as a starting point for the algorithm 1. The numerical results in Table 5.10 illustrate that the superior image has higher PSNR, and SSIM and lower MSE values in comparison to any of the degraded images individually. The plots of MSE, PSNR, and SSIM with respect to λ are shown in Figures 5.42, 5.43, and 5.44. The visual result is presented in Figure 5.45.

| (a) | (b) | (c) | (d) | (e) |

| (f) | (g) | (h) | (i) | (j) |
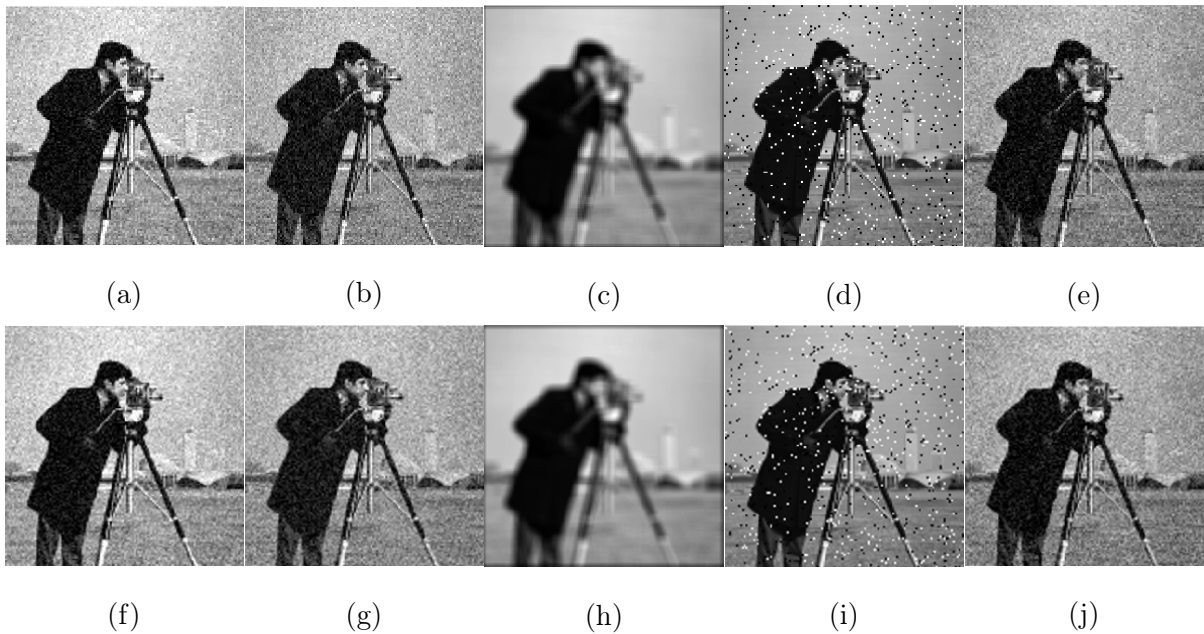
Figure 5.41: Visual results of resizing a set of distortion of the test image Cameraman by bicubic interpolation. (a) Mean contrast stretched. (b) Mean shift. (c) Blurring. (d) Salt and pepper noise. (e) Gaussian noise. (f) The resized version of mean contrast stretched. (g) The resized version of mean shift (h)The resized version of blurring. (i) The resized version of salt and pepper. (j) The resized version of Gaussian noise.

	MSE	PSNR	SSIM
Resized version of mean contrast stretched image	1043.53	17.94	0.45
Resized version of mean shift image	459.57	21.50	0.42
Resized version of Blurred image	695.54	19.70	0.56
Resized version of salt and pepper noisy image	745.85	19.40	0.55
Resized version of Gaussian noisy image	348.40	22.70	0.44
Fused image-$\lambda = 0$	459.57	21.50	0.42
Fused image-$\lambda = 4.5$	**341.61**	**22.79**	0.69
Fused image-$\lambda = 5.5$	343.93	22.76	0.71
Fused image-$\lambda = 7$	352.43	22.60	**0.72**

Table 5.10: Numerical results for the fusion of a set of distortion images. (In the case of bicubic interpolation used to resize the Cameraman images).
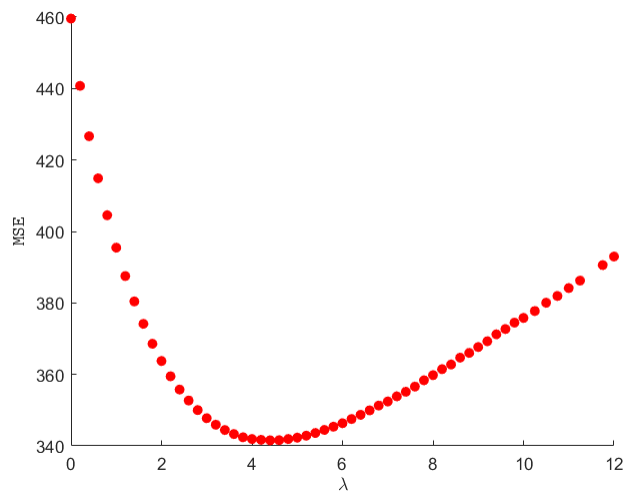
Figure 5.42: Change of MSE value with respect to λ. (In the case of bicubic interpolation used to resize the Cameraman images).
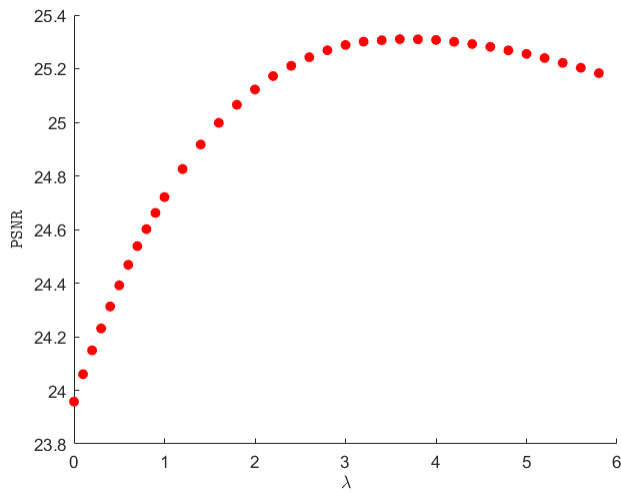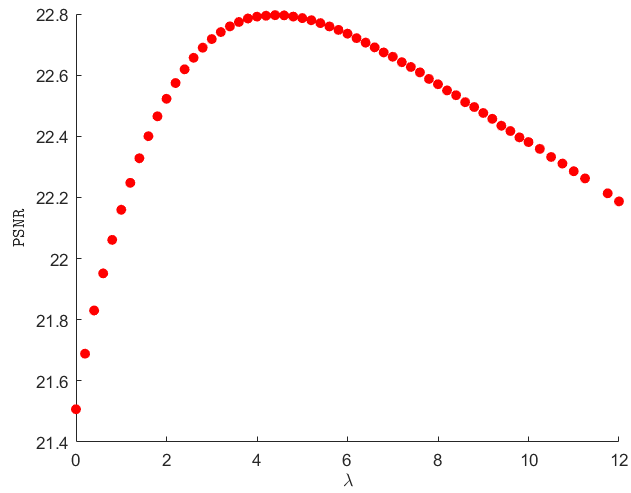
Figure 5.43: Change of PSNR value with respect to λ. (In the case of bicubic interpolation used to resize the Cameraman images).
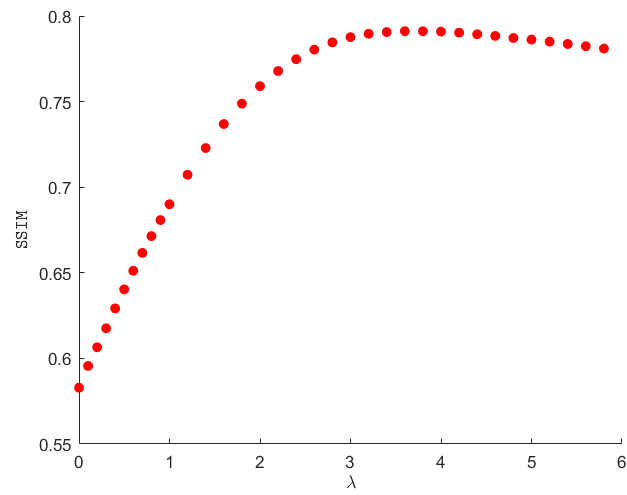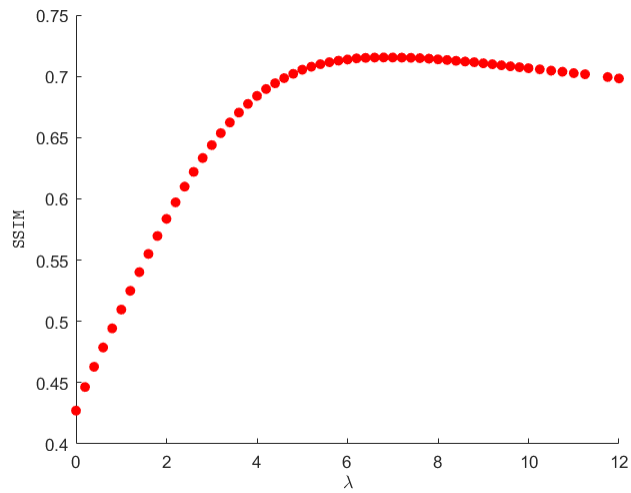

Figure 5.44: Change of SSIM value with respect to λ. (In the case of bicubic interpolation used to resize the Cameraman images).

(a)   (b)   (c)

(d)   (e)   (f)
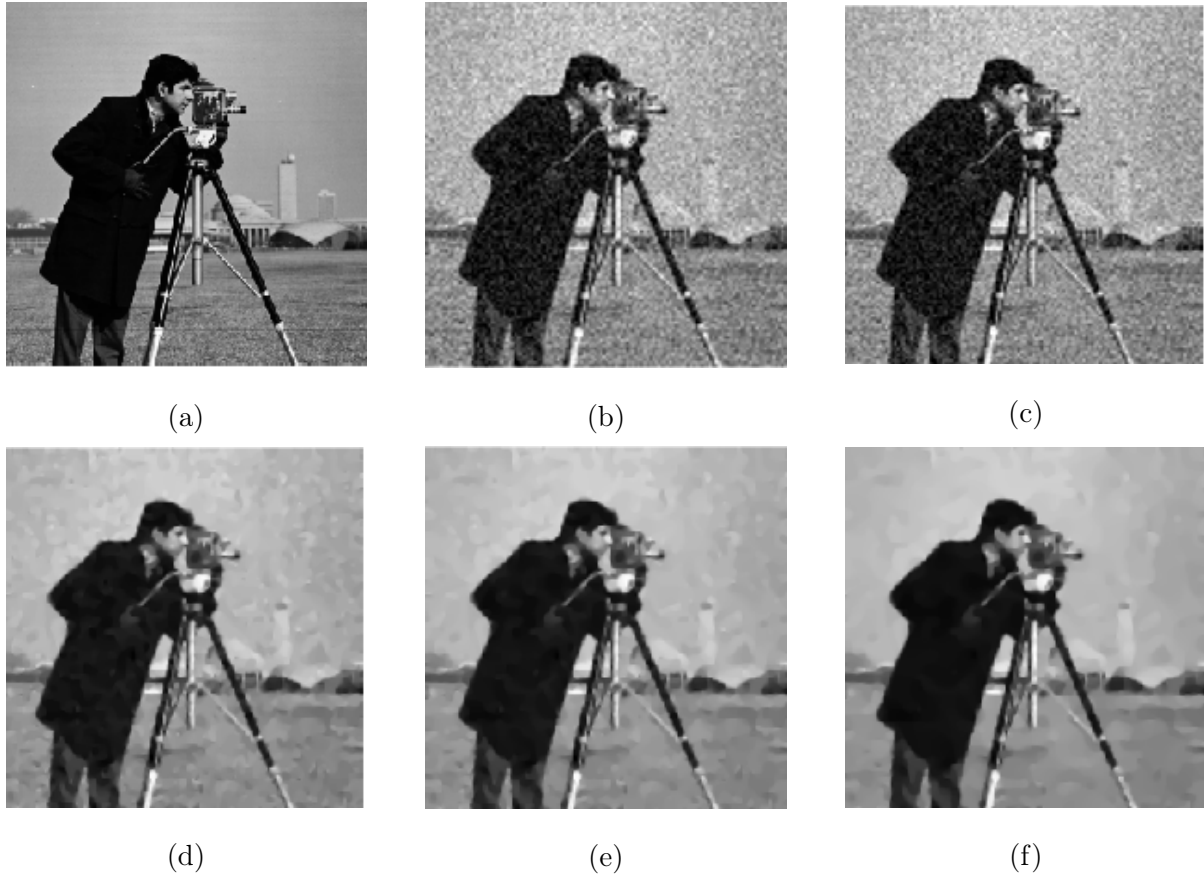
Figure 5.45: Visual results for the fusion of a set of distortion of the test image Cameraman. (a) Original image. (b) The resized version of noisy mean shift (starting point) (c) Fused image ($\lambda = 0$) (d) Fused image ($\lambda = 4.5$). (e) Fused image ($\lambda = 5.5$). (f) Fused image ($\lambda = 7$). (In the case of bicubic interpolation used to resize the Cameraman images).

2. **Result obtained by using nearest interpolation to resize the images**

   The nearest interpolation is used instead of bicubic interpolation in this experiment. The assumptions and values remain the same as those in the previous example. Table 5.11 and Figures 5.46, 5.47, 5.48, and 5.49, and 5.50 show the result both numerically and visually. It can be seen that in the Table and Figures the superior fused image was obtained when $\lambda = 9$ in terms of PSNR and SSIM in the selected range of λ.
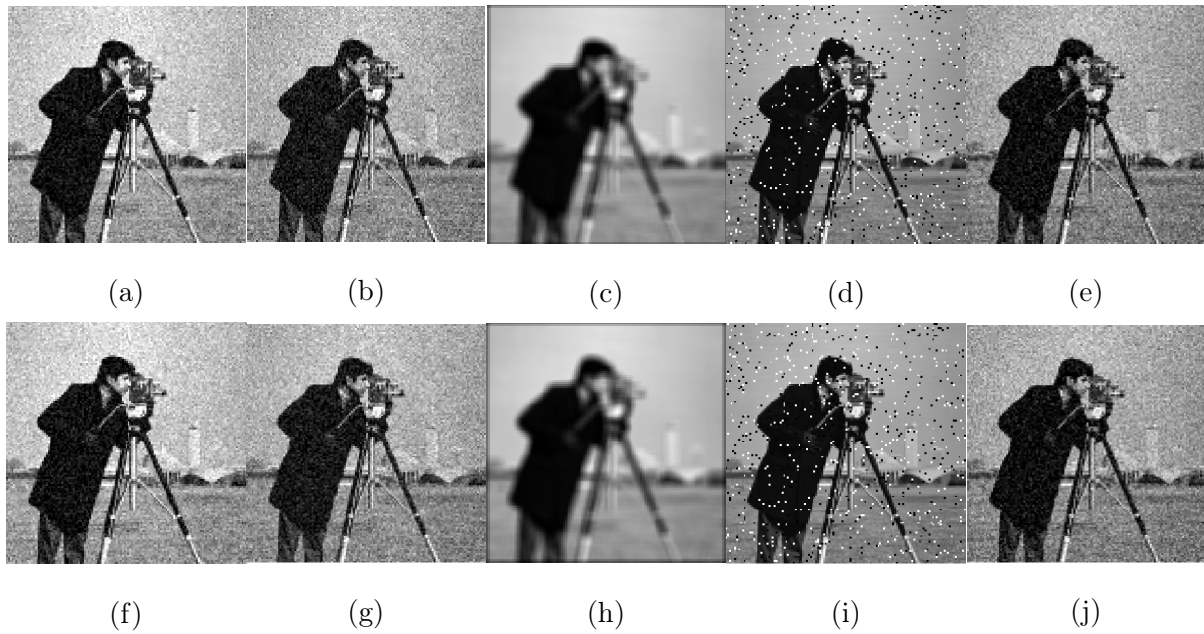
Figure 5.46: Visual results of resizing a set of distortion types of the test image Camera-man by nearest interpolation. (a) Mean contrast stretched. (b) Mean shift. (c) Blurring. (d) Salt and pepper noise. (e) Gaussian noise. (f) The resized version of mean contrast stretched. (g) The resized version of mean shift (h)The resized version of blurring. (i) The resized version of salt and pepper. (j) The resized version of Gaussian noise.

	MSE	PSNR	SSIM
Resized version of mean contrast stretched image	1201.25	17.33	0.37
Resized version of mean shift image	607.20	20.29	0.35
Resized version of Blurred image	707.48	19.63	0.56
Resized version of salt and pepper noisy image	1142.06	17.55	0.51
Resize of Gaussian noisy image	485.02	21.27	0.37
Fused image-$\lambda = 0$	607.20	20.97	0.35
Fused image-$\lambda = 1$	508.55	21.06	0.41
Fused image-$\lambda = 7.5$	**337.33**	22.85	0.72
Fused image-$\lambda = 9$	340.79	**22.86**	**0.73**

Table 5.11: Numerical results for the fusion of a set of distortion images. (In the case of nearest interpolation used to resize the Cameraman images).
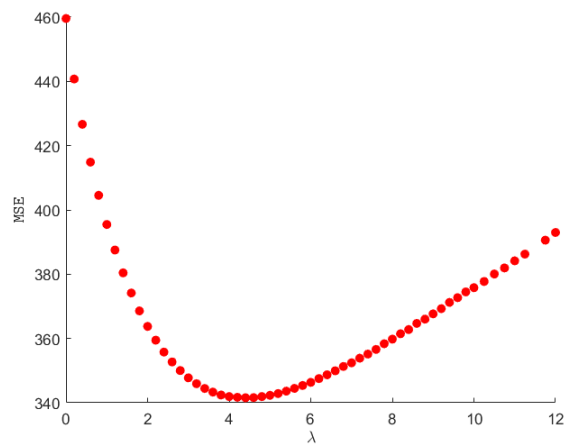
Figure 5.47: Change of MSE value with respect to λ. (In the case of nearest interpolation used to resize the Cameraman images).
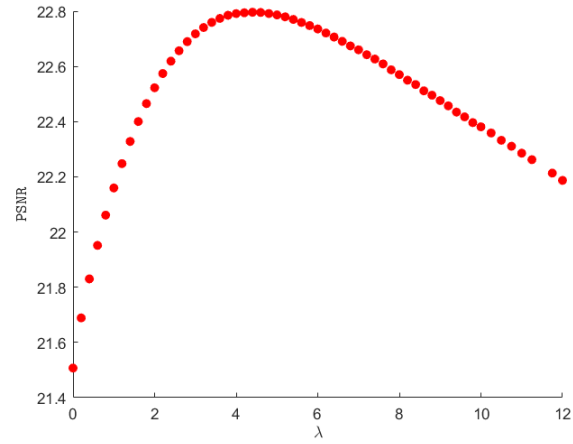
Figure 5.48: Change of PSNR value with respect to λ. (In the case of nearest interpolation used to resize the Cameraman images).
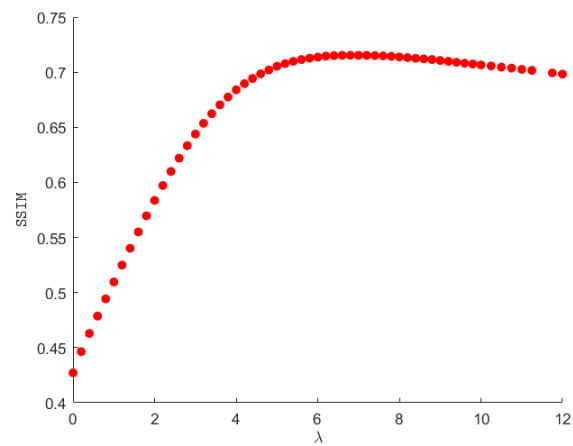
Figure 5.49: Change of SSIM value with respect to λ. (In the case of nearest interpolation used to resize the Cameraman images).
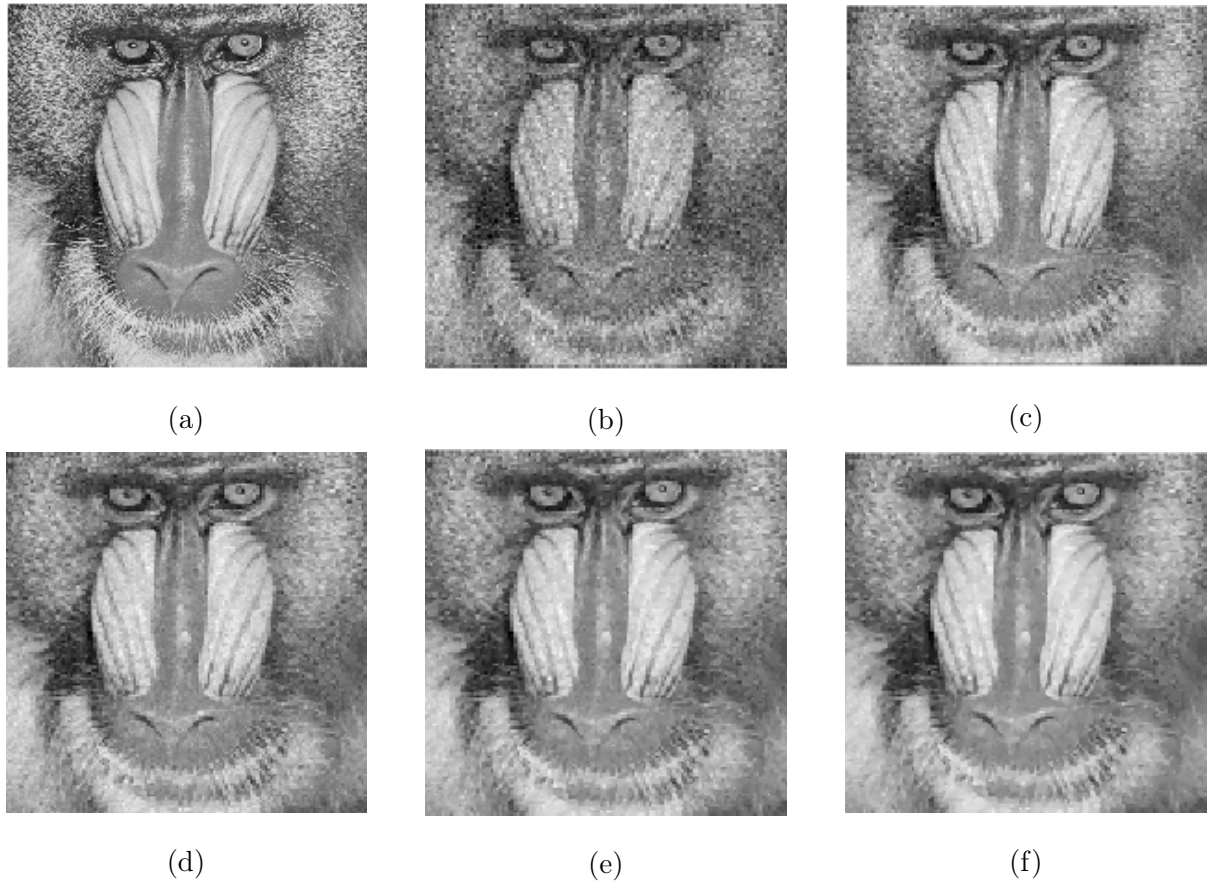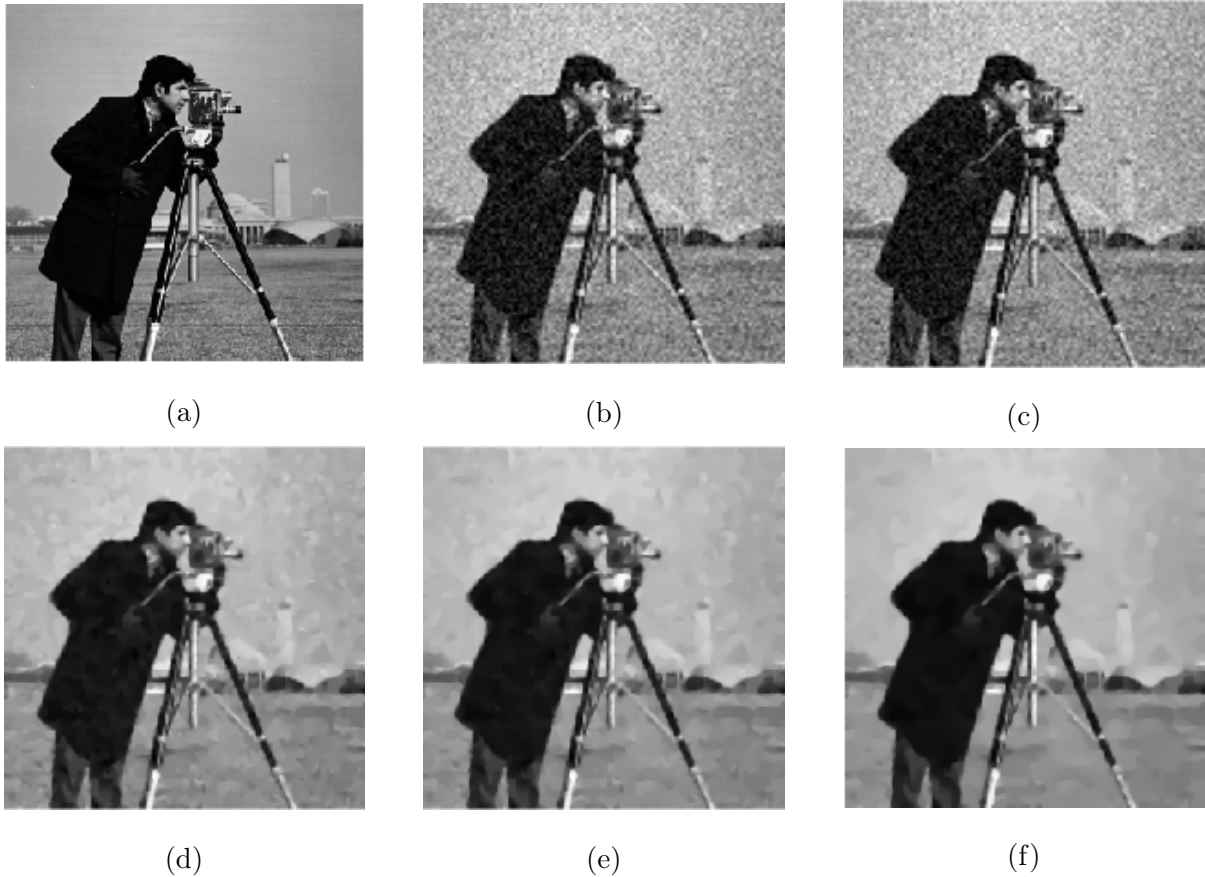
Figure 5.50: Visual results for the fusion of a set of distortion of the test image Camera-man. (a) Original image. (b) The resized version of noisy mean shifted (starting point). (c) Fused image ($\lambda = 0$). d) Fused image ($\lambda = 5.5$). (e) Fused image ($\lambda = 7.5$). (f) Fused image ($\lambda = 9.5$)(In the case of nearest interpolation used to resize the Cameraman images).

5.4.2.2   Mandrill Test Image

In order to demonstrate that the method is flexible and can be applied to a wide variety of degradation operators and images, here Mandrill image which has been degraded by local averaging followed by a downsampling factor of 4 transformation is applied. The type of degradation is the same as in the previous example. The input images need to be resized before applying the algorithm 1, as indicated in the previous example. The
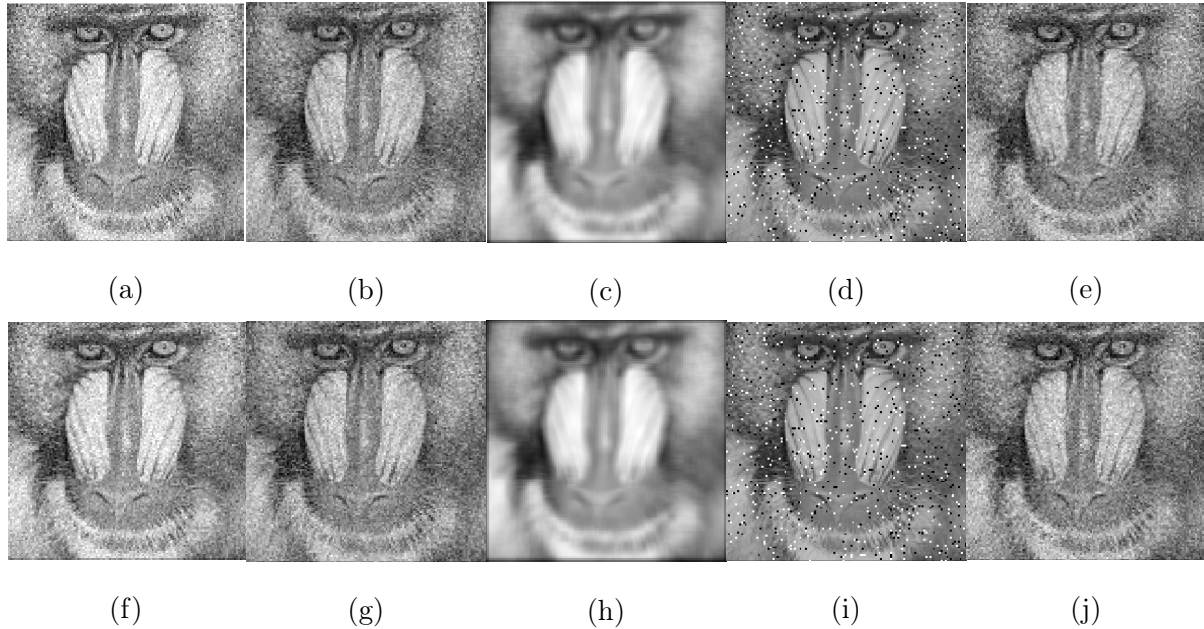
Figure 5.51: Visual results of resizing of a set of distortion of the test Mandrill image. (a) Mean contrast stretched. (b) Mean shift. (c) Blurring. (d) Salt and pepper noise. (e) Gaussian noise. (f) The resized version of mean contrast stretched. (g) The resized version of mean shift (h)The resized version of blurring. (i) The resized version of salt and pepper. (j) The resized version of Gaussian noise.

nearest interpolation is used to resize our input images (Figure 5.51). The parameters and assumptions are the same as in the previous example that the nearest interpolation is used to resize the Cameraman's images. The resized version of the Gaussian noisy image is used as the starting point. By increasing lambda, it is expected to obtain a superior fused image both virtually and numerically in terms of MSE, PSNR, and SSIM within the range of lambda selected. In Table 5.12, it is shown that a superior fused image is obtained when $\lambda = 12.5$ in the range of λ selected. Figures 5.51, 5.52, 5.53, and 5.54, and 5.55 show the result visually.

	MSE	PSNR	SSIM
Resize of mean contrast stretched image	1573.76	16.16	0.29
Resize of mean shift image	954.89	18.33	0.27
Resize of Blurred image	861.82	18.77	0.24
Resize of salt and pepper noisy image	1397.31	16.67	0.35
Resize of Gaussian noisy image	859.29	18.78	0.27
Fusion image-$\lambda = 1$	829.41	18.94	0.28
Fusion image-$\lambda = 5.5$	719.60	19.55	0.32
Fusion image-$\lambda = 9$	670.73	19.80	**0.33**
Fusion image-$\lambda = 12.5$	**668.80**	**19.87**	0.32

Table 5.12: Numerical results for the fusion of a set of distorted images. (In the case of nearest interpolation used to resize the Mandrill image).
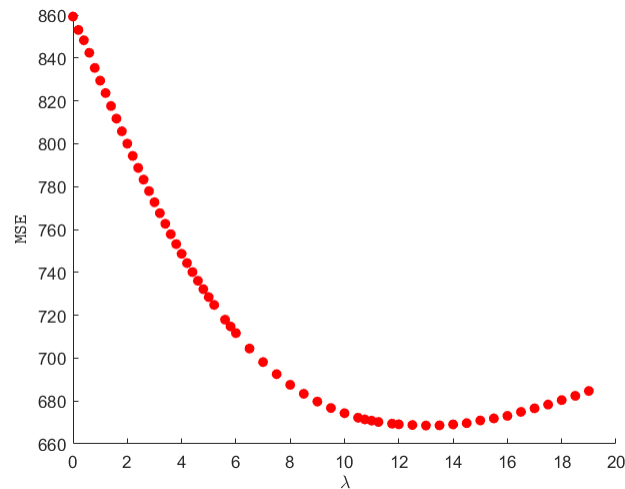
Figure 5.52: Change of MSE value with respect to λ. (In the case of nearest interpolation used to resize the Mandrill image).
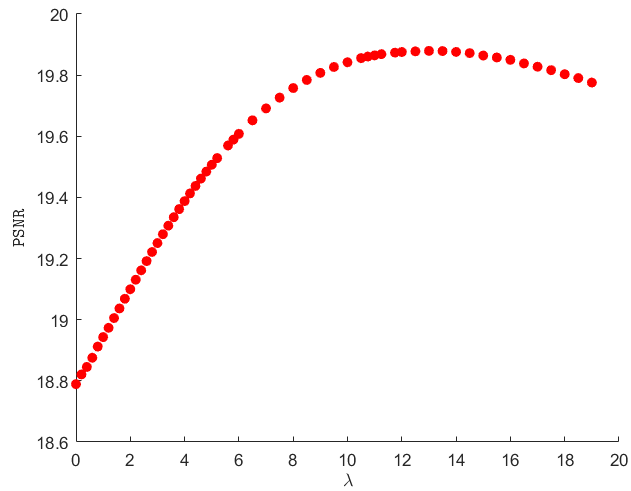
Figure 5.53: Change of PSNR value with respect to λ. (In the case of nearest interpolation used to resize the Mandrill image).
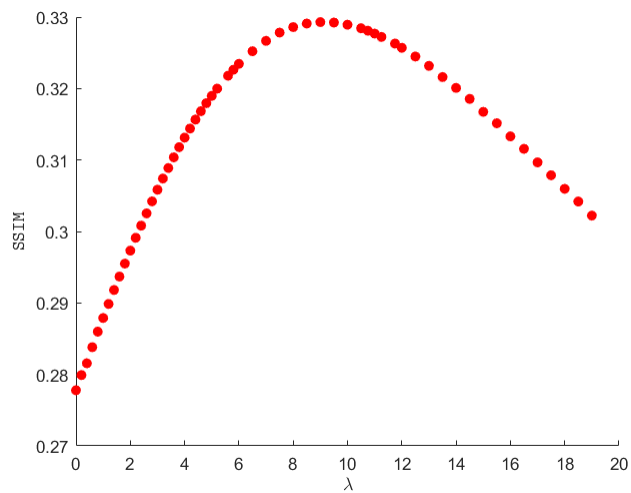
Figure 5.54: Change of SSIM value with respect to λ. (In the case of nearest interpolation used to resize the Mandrill image).

(a)   (b)   (c)

(d)   (e)   (f)
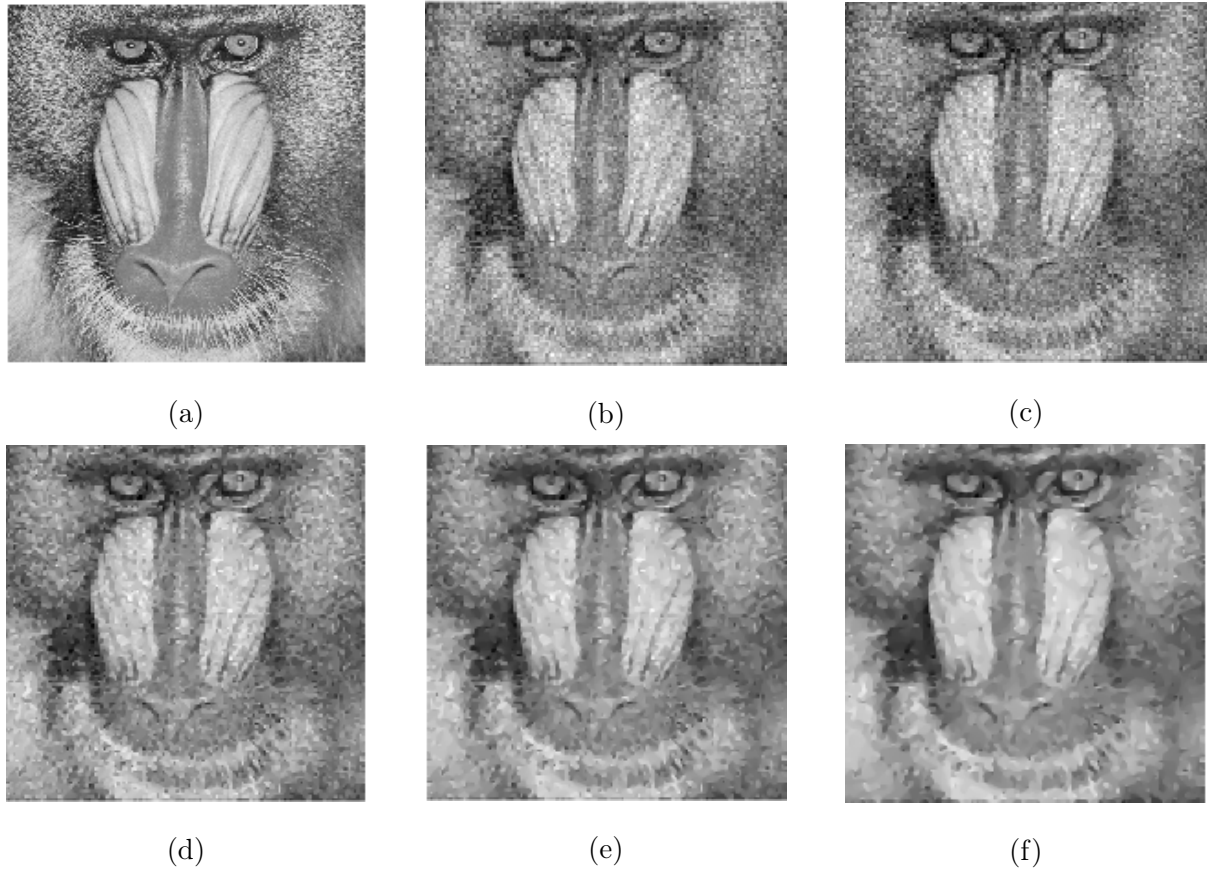
Figure 5.55: Visual results for the fusion of a set of distortion of the test image Mandrill. (a) Original image. (b) The resized version of noisy mean shift (starting point). (c) Fused image ($\lambda = 0$). d) Fused image ($\lambda = 5.5$). (e) Fused image ($\lambda = 9$). (f) Fused image ($\lambda = 12$). (In the case of nearest interpolation used to resize the Mandrill image).

Chapter 6

Conclusion and Future Work

Image fusion is primarily focused on combining information from multiple images of a scene to create a single image that conveys enhanced information. Recent years have witnessed significant advancements in image fusion methods, highlighting their progress. This thesis aims to enhance the quality of fused images through the using of the structural similarity index measure, which is widely accepted as an image quality measure consistent with the Human Visual System.

A general mathematical and computational framework for image fusion is provided, where the data fidelity term of the objective function in the inverse problem is based on SSIM-Mean. The incorporation of SSIM-Mean as a fidelity term is facilitated through the framework, allowing its application across a diverse range of applications.

To numerically solve the problem using the gradient descent method, it is necessary to compute the derivative of the SSIM, which poses a primary challenge in this context. Given the difficulty in directly differentiating the SSIM expression, an alternative expression of the SSIM in terms of mean and vector norm is introduced. This alternative formulation enables the computation of the SSIM's derivative. Subsequently, algorithms are proposed to address various SSIM-based imaging tasks, providing practical solutions for their implementation.

The obtained results demonstrate consistency with the expected results both qualitatively and quantitatively. Through the implementation of experiments, it is observed that the fused image exhibits superior performance in terms of metrics such as MSE, PSNR, SSIM, and, most importantly, visual quality when compared to the degraded input images.

A comparison between the proposed method and Brunet's methods can be made specifically for calculating SSIM Mean without considering regularization. This is due to the lack of flexibility in Brunet's method for incorporating regularization. It is observed that the numerical solution obtained from the proposed methods matches exactly with the solution obtained from Brunet's method in one dimensional case.

6.1   Thesis Contribution Highlights

The main contribution of this thesis can be summarized as follows:

- **New expression of SSIM and its derivative**

  Applying the SSIM as the data fidelity of the objective function in inverse problems poses challenges due to the inclusion of mean, variance, and covariance terms in its definition. In numerical algorithms, particularly those based on gradients, it is necessary to compute these statistical quantities and their derivatives at each iteration. In order to simplify the calculation of SSIM and its derivatives, new expressions are introduced in the background chapter, formulated in terms of mean and norm. These expressions greatly facilitate the calculation of SSIM and its derivatives, simplifying the overall process in numerical algorithms. By using this new expression, SSIM can be applied as data fidelity in the formulation of inverse problems without any additional assumptions. Previous studies, such as [61] and [60], have simplified the SSIM formula for zero-mean images and applied it to

optimization problems.

- **General formulation of SSIM-Mean optimization problems**

  The SSIM Mean optimization problems are formulated as follows:

  $$\sum_{i=1}^{n} \mathbf{p}_i \ \text{SSIM}(\mathcal{H}_i\mathbf{x}, \mathbf{y}_i) + \lambda g(\mathbf{x}),$$

  where $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ is a set of n vectors which their size is $\mathtt{n}\times 1$ and $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ is a set of associated weights across patches, $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_n\}$ is a set of degradation operators which its size is $\mathtt{n} \times \mathtt{m}$, $g(\mathbf{x})$ is a regularization term, and λ is its corresponding regularization parameter.

  As indicated above, $\mathcal{H}_i$ is considered in a generalized form, i.e. it can be regarded as any degradation operator. Based on our knowledge all previous works consider $\mathcal{H} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. The proposed formulation offers flexibility for incorporating additional terms into the criteria and applying various types of regularization.

6.2   Future Work

The proposed methods in this thesis open several new directions for future work.

- Other types of distortions can be considered in the future as a result of this general formulation. For instance, nighttime images pose a challenge in terms of comprehensibility due to the lack of background context caused by poor illumination. For instance, when observing an image captured by a traffic camera on the web or displayed on TV, it becomes challenging to determine the specific location within the town, the number of lanes on the highway, or nearby buildings. The screen merely displays moving headlights. To address this issue, our solution is based on a simple observation. The fact that the scene can be observed by the traffic camera all day

long and a high-quality background can be created is exploited. Subsequently, the context of the low-quality image can be enhanced by fusing the appropriate pixels.

- The formulation of the fusion image allows for the application of other degradation operators and regularization methods. In terms of changing the degradation operator, one example can be applying a local averaging operator. The result will be affected by changing the degradation operator, as gradient-based methods are used and consequently the derivative also includes the degradation operator. In terms of regularization, changing the regularization term would have an impact on the final result as it is a part of the derivative of the objective function. Total variation regularization is used here and it perceives the edges, but an alternative choice could be L2 regularization, which is differentiable and can cause considerable smoothing of the solution [19].

- The application of this method extends to various image processing domains, including medical imaging. Medical imaging techniques, including Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Single-Photon Emission Computed Tomography (SPECT), provide clinicians with valuable insights into the structural characteristics and soft tissue of the human body. Each imaging method exhibits distinct characteristics, and different sensors capture unique information about the same anatomical area. The objective of image fusion is to enhance contrast, fusion quality, and the overall perceptual experience. The fused image should meet to the following criteria: (a) it should faithfully preserve all information from the source images, (b) it should avoid introducing any artificial information such as artifacts, and (c) undesirable states such as noise should be minimized or eliminated. The proposed methods have the potential to be applied in the field of medicine to obtain superior fused images.

Appendix A

Calculation of SSIM Derivative

A.1 Matrix Calculus

This section defines some functions and calculates their derivatives to use them for our purpose function's derivative.

1. Let

$$\Theta(\mathbf{x}) = \mathbf{c}^T \mathbf{x}, \tag{A.1}$$

where $\mathbf{c}$ is a constant vector and $\mathbf{x} = [x_1, \ldots, x_n]$ is a vector. $\Theta(\mathbf{x})$ can be expressed as

$$\Theta(\mathbf{x}) = \sum_{j=1}^{n} c_j x_j. \tag{A.2}$$

Derivative of the equation A.2 is

$$\frac{\partial \Theta(\mathbf{x})}{\partial x_k} = \sum_{j=1}^{n} c_j \frac{\partial x_j}{\partial x_k} = \sum_{j=1}^{n} x_j \delta_{jk} = x_k. \tag{A.3}$$

$$\nabla \Theta(\mathbf{x}) = \mathbf{c}. \tag{A.4}$$

2. $\psi(\mathbf{x})$ is defined as follows,

$$\psi(\mathbf{x}) = \mathbf{x}^T \mathcal{B} \mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i b_{ij} x_j. \tag{A.5}$$

The derivative of $\psi(\mathbf{x})$ is

$$\frac{\partial \psi(\mathbf{x})}{\partial \mathsf{x}_k} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathsf{b}_{ij}\frac{\partial(\mathsf{x}_i\mathsf{x}_j)}{\partial \mathsf{x}_k}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathsf{b}_{ij}(\delta_{ik}x_j + \mathsf{x}_i\delta_{jk})$$

$$= \sum_{j=1}^{n} \mathsf{b}_{kj}\mathsf{x}_j + \sum_{i=1}^{n} \mathsf{b}_{ik}\mathsf{x}_i \qquad \text{(A.6)}$$

$$= (\mathcal{B}x)_k + \sum_{i=1}^{n}(\mathcal{B}^T)_{ki}\mathsf{x}_i$$

$$= (\mathcal{B}\mathbf{x})_k + (\mathcal{B}^T\mathbf{x})_k.$$

As a result, it can be written as,

$$\nabla \psi(\mathbf{x}) = (\mathcal{B} + \mathcal{B}^T)\mathbf{x}. \qquad \text{(A.7)}$$

3. The derivative of $\|\phi\mathbf{x} - \mathbf{x}\|_2^2$ is aimed to calculated.

   Using

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}} \qquad \text{(A.8)}$$

It can be obtained using the transpose operator's properties,

$$\|\phi\mathbf{x} - \mathbf{y}\|_2^2 = (\phi\mathbf{x} - \mathbf{y})^T(\phi\mathbf{x} - \mathbf{y})$$

$$= ((\phi\mathbf{x})^T - \mathbf{y}^T)(\phi\mathbf{x} - \mathbf{y}) = (\mathbf{x}^T\phi^T - \mathbf{y}^T)(\phi\mathbf{x} - \mathbf{y})$$

$$= \mathbf{x}^T\phi^T\phi\mathbf{x} - (\phi\mathbf{x})^T\mathbf{y} - \mathbf{y}^T\phi\mathbf{x} + \mathbf{y}^T\mathbf{y} \qquad \text{(A.9)}$$

$$= \mathbf{x}^T\phi^T\phi\mathbf{x} - 2\mathbf{y}^T\phi\mathbf{x} + \mathbf{y}^T\mathbf{y}$$

$$= \mathbf{x}^T\phi^T\phi\mathbf{x} - 2(\phi^T\mathbf{y})^T\mathbf{x} + \mathbf{y}^T\mathbf{y}$$

The formulas A.1, A.4, A.5, and A.7, can be used

$$\nabla(\|\phi\mathbf{x} - \mathbf{y}\|_2^2) = -2\phi^T\mathbf{y} + (\phi^T\phi + (\phi^T\phi)^T)\mathbf{x}, \qquad \text{(A.10)}$$

,

$$\nabla(\|\phi\mathbf{x} - \mathbf{y}\|_2^2) = -2\phi^T\mathbf{y} + 2\phi^T\phi\mathbf{x} = 2(\phi^T\phi\mathbf{x} - \phi^T\mathbf{y}). \qquad \text{(A.11)}$$

Note,

$$(\phi^T\phi)^T = (\phi)^T(\phi^T)^T = \phi^T\phi. \qquad \text{(A.12)}$$

A.2   Derivatives of Matrices and Vectors

Let $\mathbf{y}$ be a column vector $\mathtt{m} \times 1$ and $\mathcal{H}$ be a $\mathtt{n} \times \mathtt{m}$ matrix. In this case, $\mathcal{H}\mathbf{y}$ is a vector of size $\mathtt{n} \times 1$.

$\overline{\mathcal{H}\mathbf{x}}$ is assumed to be the mean of $\mathcal{H}\mathbf{x}$ (it is scalar) and its size is 1×1.

$\overline{\mathcal{H}\mathbf{y}}$ is defined as

$$\overline{\mathcal{H}\mathbf{x}} = \frac{1}{n}\left[1 \cdots 1\right]() = \frac{1}{\mathtt{n}}\mathbf{1}\mathcal{H}\mathbf{x}, \tag{A.13}$$

where $\mathbf{1} = \left[1 \cdots 1\right]$ is a $1 \times \mathtt{n}$ vector and $\mathcal{H}\mathbf{y}$ is $\mathtt{n} \times 1$. The derivative of equation A.13 is

$$\frac{\partial \overline{\mathcal{H}\mathbf{y}}}{\partial \mathbf{y}} = \frac{1}{\mathtt{n}}\mathbf{1}\mathcal{H}. \tag{A.14}$$

It can be written as follows,

$$\frac{\partial \overline{\mathcal{H}\mathbf{y}}}{\partial y} = \frac{1}{\mathtt{n}}(\mathbf{1}\mathcal{H})^T = \frac{1}{\mathtt{n}}\mathcal{H}^T\mathbf{1}^T. \tag{A.15}$$

Due to the fact that $\overline{\mathcal{H}\mathbf{y}}$ is a scalar, the following can be concluded

$$\overline{(\mathcal{H}\mathbf{y})^T} = \overline{\mathcal{H}\mathbf{y}} \Rightarrow \frac{\partial \overline{(\mathcal{H}\mathbf{y})^T}}{\partial \mathbf{y}} = \frac{\partial \overline{\mathcal{H}\mathbf{y}}}{\partial \mathbf{y}} = \frac{1}{\mathtt{n}}\mathbf{1}\mathcal{H}, \tag{A.16}$$

and,

$$\overline{(\mathcal{H}\mathbf{y})^T} = \overline{\mathcal{H}\mathbf{y}} \Rightarrow \frac{\partial \overline{(\mathcal{H}\mathbf{y})^T}}{\partial \mathbf{y}} = \frac{\partial \overline{\mathcal{H}\mathbf{y}}}{\partial \mathbf{y}} = \frac{1}{\mathtt{n}}\mathcal{H}^T\mathbf{1}^T. \tag{A.17}$$

$\left\|\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{x}\right\|_2^2$ is calculated. The Size of $\mathbf{1}^T$, $\mathcal{H}\mathbf{y}$ and $\mathbf{x}$ is $\mathtt{n}\times1$. $\left\|\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{x}\right\|_2^2$ is calculated as follows,

$$\begin{aligned}
\left\|\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + x\right\|_2^2 &= (\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{x})^T(\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{x}) \\
&= ((\mathcal{H}\mathbf{y})^T - (\mathbf{1}^T\overline{\mathcal{H}\mathbf{y}})^T + \mathbf{x}^T)(\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{x}) \\
&= (\mathbf{y}^T\mathcal{H}^T - (\overline{\mathcal{H}\mathbf{y}})^T\mathbf{1} + \mathbf{x}^T)(\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{x}) \\
&= (\mathbf{y}^T\mathcal{H}^T - (\overline{\mathcal{H}\mathbf{y}})\mathbf{1} + \mathbf{x}^T)(\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{x}) \\
&= \mathbf{y}^T\mathcal{H}^T\mathcal{H}\mathbf{y} - \mathbf{y}^T\mathcal{H}^T\mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{y}^T\mathcal{H}^T\mathbf{x} - (\overline{\mathcal{H}\mathbf{y}})\mathbf{1}\mathcal{H}\mathbf{y} + (\overline{\mathcal{H}\mathbf{y}})\mathbf{1}\mathbf{1}^T(\overline{\mathcal{H}\mathbf{y}}) \\
&\quad - (\overline{\mathcal{H}\mathbf{y}})\mathbf{1}\mathbf{x} + \mathbf{x}^T\mathcal{H}\mathbf{y} - \mathbf{x}^T\mathbf{1}^T(\overline{\mathcal{H}\mathbf{y}}) + \mathbf{x}^T\mathbf{x}.
\end{aligned}$$

$$\tag{A.18}$$

$\mathcal{H}\mathbf{y}$ is a vector, and so is $\mathbf{1}^T\overline{\mathcal{H}\mathbf{y}}$. So when we multiply one by another, it doesn't matter what the order is (as long as the dimensions work out). Also this is correct for $\mathbf{y}^T\mathcal{H}^T\mathbf{x}$, $\mathbf{x}^T\mathbf{y}\mathcal{H}$, $(\overline{\mathcal{H}\mathbf{y}})\mathbf{1}\mathbf{x}$, and $\mathbf{x}^T\mathbf{1}^T(\overline{\mathcal{H}\mathbf{y}})$.

As shown in the following, $\mathbf{1}\mathbf{1}^T = n$ and $\mathbf{1}^T\mathbf{1} = J_{\mathrm{n}\times\mathrm{n}}$

$$\mathbf{1}\mathbf{1}^T = \begin{bmatrix} 1 \cdots 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = 1 + \cdots + 1 = \mathrm{n}. \tag{A.19}$$

$$\mathbf{1}^T\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 \cdots 1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \mathbf{J}_{\mathrm{n}\times\mathrm{n}}. \tag{A.20}$$

The equation A.18 is simplified by substituting equation A.13.

$$\begin{aligned} \left\|\mathcal{H}\mathbf{y} - \mathbf{1}^T\overline{\mathcal{H}\mathbf{y}} + \mathbf{x}\right\|_2^2 &= \mathbf{y}^T\mathcal{H}^T\mathcal{H}\mathbf{y} - 2(\overline{\mathcal{H}\mathbf{y}})\mathbf{1}\mathcal{H}\mathbf{y} + 2\mathbf{x}^T\mathcal{H}\mathbf{y} - 2(\overline{\mathcal{H}\mathbf{y}})\mathbf{1}\mathbf{x} + \mathrm{n}(\overline{\mathcal{H}\mathbf{y}})^2 + \mathbf{x}^T\mathbf{x} \\ &= \mathbf{y}^T\mathcal{H}^T\mathcal{H}\mathbf{y} - 2(\frac{1}{\mathrm{n}}\mathbf{1}\mathcal{H}\mathbf{y})\mathbf{1}\mathcal{H}\mathbf{y} + 2\mathbf{x}^T\mathcal{H}\mathbf{y} - 2(\frac{1}{\mathrm{n}}\mathbf{1}\mathcal{H}\mathbf{y})\mathbf{1}x + \mathrm{n}(\frac{1}{\mathrm{n}}\mathbf{1}\mathcal{H}\mathbf{y})^2 + \mathbf{x}^T\mathbf{x} \\ &= \mathbf{y}^T\mathcal{H}^T\mathcal{H}\mathbf{y} - \frac{2}{n}(\mathbf{1}\mathcal{H}\mathbf{y})^2 + 2\mathbf{x}^T\mathcal{H}\mathbf{y} - \frac{2}{\mathrm{n}}(\mathbf{1}\mathcal{H}\mathbf{y})\mathbf{1}\mathbf{x} + \frac{\mathrm{n}}{\mathrm{n}^2}(\mathbf{1}\mathcal{H}\mathbf{y})^2 + \mathbf{x}^T\mathbf{x} \\ &= \mathbf{y}^T\mathcal{H}^T\mathcal{H}\mathbf{y} - \frac{1}{\mathrm{n}}(\mathbf{1}\mathcal{H}\mathbf{y})^2 + 2\mathbf{x}^T\mathcal{H}\mathbf{y} - \frac{2}{\mathrm{n}}(\mathbf{1}\mathcal{H}\mathbf{y})\mathbf{1}\mathbf{x} + \mathbf{x}^T\mathbf{x}. \end{aligned} \tag{A.21}$$

In terms of the size of each term of the A.21,

$$\begin{aligned} &\mathbf{y}^T : 1 \times \mathrm{m} \quad \mathcal{H}^T : \mathrm{m} \times \mathrm{n} \quad \mathcal{H} : \mathrm{n} \times \mathrm{m} \quad \mathbf{y} : \mathrm{m} \times 1 \longrightarrow \mathbf{y}^T\mathcal{H}^T\mathcal{H}\mathbf{y} : 1 \times 1, \\ &\mathbf{1} : 1 \times \mathrm{n} \quad \mathcal{H}\mathbf{y} : \mathrm{n} \times 1 \longrightarrow (\mathbf{1}\mathcal{H}\mathbf{y})^2 : 1 \times 1, \\ &\mathbf{x}^T : 1 \times \mathrm{n} \quad \mathcal{H}\mathbf{y} : \mathrm{n} \times 1 \longrightarrow \mathbf{x}^T\mathcal{H}\mathbf{y} : 1 \times 1, \\ &\mathbf{1} : 1 \times \mathrm{n} \quad \mathcal{H}\mathbf{y} : \mathrm{n} \times 1 \quad \mathbf{1} : 1 \times \mathrm{n} \quad \mathbf{x} : \mathrm{n} \times 1 \longrightarrow (\mathbf{1}\mathcal{H}\mathbf{y})\mathbf{1}\mathbf{x} : 1 \times 1, \\ &\mathbf{x}^T : 1 \times \mathrm{n} \quad \mathbf{x} : \mathrm{n} \times 1 \longrightarrow \mathbf{x}^T\mathbf{x} : 1 \times 1. \end{aligned} \tag{A.22}$$

Note that $\mathbf{x}^T\mathcal{H}\mathbf{y}$ is just the dot product between $\mathcal{H}\mathbf{y}$ and $\mathbf{x}$ regarded as a column vector, and respectively for other ones.

Using formulas A.13 and A.20, the equation A.21 can be rewritten

$$
\begin{aligned}
\left\| \mathcal{H}\mathbf{y} - \mathbf{1}^T \overline{\mathcal{H}\mathbf{y}} + \mathbf{x} \right\|_2^2 &= \left\| \mathcal{H}\mathbf{y} - \mathbf{1}^T \frac{1}{n}\mathbf{1}\mathcal{H}\mathbf{y} + \mathbf{x} \right\|_2^2 \\
&= \left\| \mathbf{I}\mathcal{H}\mathbf{y} - \frac{1}{n}\mathbf{1}^T \mathbf{1}\mathcal{H}\mathbf{y} + \mathbf{x} \right\|_2^2 \\
&= \left\| \mathbf{I}\mathcal{H}\mathbf{y} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}}\mathcal{H}\mathbf{y} + \mathbf{x} \right\|_2^2 \qquad (A.23) \\
&= \left\| I\mathcal{H}\mathbf{y} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}}\mathcal{H}\mathbf{y} + \mathbf{x} \right\|_2^2 \\
&= \left\| (\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H}\mathbf{y} + \mathbf{x} \right\|_2^2 ,
\end{aligned}
$$

where $\mathbf{I}$ is the identity matrix and its size is $\mathbf{n}\times\mathbf{n}$.

Regarding the size of the terms of A.23

$$
\begin{aligned}
\mathbf{I} : \mathbf{n}\times\mathbf{n} \quad \mathbf{J} : \mathbf{n}\times\mathbf{n} \quad &\Rightarrow (\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}}) : \mathbf{n}\times\mathbf{n}, \\
(\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}}) : \mathbf{n}\times\mathbf{n} \quad \mathcal{H} : \mathbf{n}\times\mathbf{m} \quad \mathbf{y} : \mathbf{m}\times 1 &\Rightarrow (\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H}\mathbf{y} : \mathbf{n}\times 1, \qquad (A.24) \\
\mathbf{x} : \mathbf{n}\times 1. &
\end{aligned}
$$

A derivative of A.23 can be taken by using A.11.

$$
\begin{aligned}
\nabla(\left\| \mathcal{H}\mathbf{y} - \mathbf{1}^T \overline{\mathcal{H}\mathbf{y}} + \mathbf{x} \right\|_2^2) &= \nabla(\left\| (\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H}\mathbf{y} + \mathbf{x} \right\|_2^2) \\
&= 2\left(\left(((\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H})^T ((\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H}) \right)\mathbf{y} + ((\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H})^T\mathbf{x} \right) \\
&= 2\left(\left((\mathcal{H}^T(\mathbf{I}^T - \frac{1}{n}\mathbf{J}^T_{\mathbf{n}\times\mathbf{n}})) ((\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H}) \right)\mathbf{y} + (\mathcal{H}^T(\mathbf{I}^T - \frac{1}{n}\mathbf{J}^T_{\mathbf{n}\times\mathbf{n}}))\mathbf{x} \right) \\
&= 2\left(\left((\mathcal{H}^T(\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})) ((\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H}) \right)\mathbf{y} + (\mathcal{H}^T(\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}}))\mathbf{x} \right) \\
&= 2\left(\left((\mathcal{H}^T(\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})^2\mathcal{H}) \right)\mathbf{y} + (\mathcal{H}^T(\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}}))\mathbf{x} \right)
\end{aligned}
$$

$$(A.25)$$

The size of each term in equation A.25 is as follows

$$
\begin{aligned}
&\mathcal{H}^T : \mathtt{m \times n} \quad (\mathbf{I} - \frac{1}{\mathtt{n}}\mathbf{J_{n\times n}}) : \mathtt{n \times n} \quad \mathcal{H}; nn \times \mathtt{m} \Rightarrow \left(\left(\mathcal{H}^T(\mathbf{I} - \frac{1}{\mathtt{n}}\mathbf{J_{n\times n}})^2 \mathcal{H} \right) \right) : \mathtt{m \times m} \\
&\left(\left(\mathcal{H}^T(\mathbf{I} - \frac{1}{\mathtt{n}}\mathbf{J_{n\times n}})^2 \mathcal{H} \right) \right) : \mathtt{m \times m} \quad \mathbf{y} : \mathtt{m \times 1} \Rightarrow \left(\left(\mathcal{H}^T(\mathbf{I} - \frac{1}{\mathtt{n}}\mathbf{J_{n\times n}})^2 \mathcal{H} \right) \right) \mathbf{y} : \mathtt{m \times 1} \\
&\mathcal{H}^T : \mathtt{m \times n} \quad (\mathbf{I} - \frac{1}{\mathtt{n}}\mathbf{J_{n\times n}}) : \mathtt{n \times n} \quad \Rightarrow \left(\left(\mathcal{H}^T(I - \frac{1}{\mathtt{n}}\mathbf{J_{n\times n}})^2 \right) \right) : \mathtt{m \times n} \\
&\left(\left(\mathcal{H}^T(\mathbf{I} - \frac{1}{\mathtt{n}}\mathbf{J_{n\times n}})^2 \right) \right) : \mathtt{m \times n} \quad \mathbf{x} : \mathtt{n \times 1} \Rightarrow (\mathcal{H}^T(\mathbf{I} - \frac{1}{\mathtt{n}}\mathbf{J_{n\times n}}))\mathbf{x} : \mathtt{m \times 1}
\end{aligned}
\tag{A.26}
$$

A.3   Derivative of SSIM Mean

Let $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ is a set of n vectors, $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ is a set of associated weights , and $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_n\}$ a set of matrix operators. The derivative of the following formula is taken,

$$
\sum_{i=1}^{n} \mathbf{p}_i \left(\frac{2\,\overline{\mathcal{H}_i \mathbf{x}}\,\overline{\mathbf{y}_i} + \varepsilon_1}{\overline{\mathcal{H}_i \mathbf{x}}^2 + \overline{\mathbf{y}_i}^2 + \varepsilon_1} \right)
$$
$$
\left(\frac{\left\| \mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}} + \mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i} \right\|_2^2 - \left\| \mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}} \right\|_2^2 - \left\| \mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i} \right\|_2^2 + \varepsilon_2}{\left\| \mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}} \right\|_2^2 + \left\| \mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i} \right\|_2^2 + \varepsilon_2} \right).
\tag{A.27}
$$

Note that $\bar{\mathbf{x}}$ is scalar and the column vector, hence setting $\bar{x} = \mathbf{1}^T x$. The cost function is

$$
P = \sum_{i=1}^{n} \mathbf{p}_i S_1 S_2,
\tag{A.28}
$$

where,

$$
S_1 = \frac{2\,\overline{\mathcal{H}_i \mathbf{x}}\,\overline{\mathbf{y}_i} + \varepsilon_1}{\overline{\mathcal{H}_i \mathbf{x}}^2 + \overline{\mathbf{y}_i}^2 + \varepsilon_1},
\tag{A.29}
$$

$$
S_2 = \frac{\left\| \mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}} + \mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i} \right\|_2^2 - \left\| \mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}} \right\|_2^2 - \left\| \mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i} \right\|_2^2 + \varepsilon_2}{\left\| \mathcal{H}_i\mathbf{x} - \mathbf{1}^T\overline{\mathcal{H}_i\mathbf{x}} \right\|_2^2 + \left\| \mathbf{y}_i - \mathbf{1}^T\overline{\mathbf{y}_i} \right\|_2^2 + \varepsilon_2}.
\tag{A.30}
$$

Hence by using the product rule,

$$
\frac{\partial p}{\partial \mathbf{x}} = \sum_{i=1}^{n} \mathbf{p}_i \left[\frac{\partial S_1}{\partial \mathbf{x}} S_2 + \frac{\partial S_2}{\partial \mathbf{x}} S_1 \right]
\tag{A.31}
$$

According the equation A.14 and A.15,

$$
\begin{aligned}
\frac{\partial S_1}{\partial \mathbf{x}} &= \frac{2\overline{\mathbf{y}_i}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1) - 2\overline{\mathcal{H}_i\mathbf{x}}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T(2\overline{\mathbf{y}_i}\,\overline{\mathcal{H}_i\mathbf{x}} + \varepsilon_1)}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)^2} \\
&= \frac{2\overline{\mathbf{y}_i}\frac{1}{n}\mathcal{H}\mathcal{H}_i{}^T\mathbf{1}^T(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1) - 2\overline{\mathcal{H}_i\mathbf{x}}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T(2\overline{\mathbf{y}_i}\,\overline{\mathcal{H}_i\mathbf{x}} + \varepsilon_1)}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)} \\
&= \frac{2\overline{\mathbf{y}_i}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)} - \frac{2\overline{\mathcal{H}_i\mathbf{x}}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T(2\overline{\mathbf{y}_i}\,\overline{\mathcal{H}_i\mathbf{x}} + \varepsilon_1)}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)} \\
&= \frac{2\overline{\mathbf{y}_i}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)} - \frac{2\overline{\mathcal{H}_i\mathbf{x}}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T S_1}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)} \\
&= \frac{2\overline{\mathbf{y}_i}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T - 2\overline{\mathcal{H}_i\mathbf{x}}\frac{1}{n}\mathcal{H}_i{}^T\mathbf{1}^T S_1}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)} \\
&= \frac{\frac{2}{n}\mathcal{H}_i{}^T\mathbf{1}^T(\overline{\mathbf{y}_i} - \overline{\mathcal{H}_i\mathbf{x}}S_1)}{(\overline{\mathbf{y}_i}^2 + \overline{\mathcal{H}_i\mathbf{x}}^2 + \varepsilon_1)}
\end{aligned}
$$

(A.32)

In regards to the size of each term in A.32

$$\mathcal{H}_i{}^T : \mathtt{m} \times \mathtt{n} \quad \mathbf{1}^T : \mathtt{n} \times 1 \quad \overline{\mathbf{y}_i} : 1 \times 1 \quad \overline{\mathcal{H}_i\mathbf{x}} : 1 \times 1 \quad S_1 : 1 \times 1 \longrightarrow \mathcal{H}_i{}^T\mathbf{1}^T(\overline{\mathbf{y}_i} - \overline{\mathcal{H}_i\mathbf{x}}S_1) : \mathtt{m} \times 1$$

$$\overline{\mathbf{y}_i}^2 : 1 \times 1 \quad \overline{\mathcal{H}_i\mathbf{x}}^2 : 1 \times 1 \quad \varepsilon_1 : 1 \times 1$$

(A.33)

By considering the equation A.25

$$
\begin{aligned}
\frac{\partial S_2}{\partial \mathbf{x}} &= \Bigg[2\Bigg(\Big(\big(\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{I}_{n\times n})^2\mathcal{H}_i)\big)\Big)\mathbf{x} + (\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n}))(\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}) \Bigg) \\
&\quad - 2\Big(\big(\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})^2\mathcal{H}_i)\big)\Big)\mathbf{x} \Bigg] \frac{\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2}{(\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2)^2} \\
&\quad - \frac{2\Big(\big(\mathcal{H}_i^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})^2\mathcal{H}_i)\big)\Big)\mathbf{x}\Big(\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i} + \mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{H_i x}\right\|_2^2\Big)}{(\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2)^2} \\
&\quad - \frac{2\Big(\big(\mathcal{H}_i^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})^2\mathcal{H}_i)\big)\Big)\mathbf{x}\Big(-\left\|\mathbf{y}_i - \mathbf{1}^T\,\bar{\mathbf{y}}_i\right\|_2^2 - \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2\Big)}{(\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2)^2} \\
&= \frac{2\Big(\big(\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})^2\mathcal{H}_i)\big)\Big)\mathbf{x} + 2(\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n}))(\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}) - 2\Big(\big(\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})^2\mathcal{H}_i)\big)\Big)\mathbf{x}}{(\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2)} \\
&\quad - \frac{2\Big(\big(\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})^2\mathcal{H}_i)\big)\Big)\mathbf{x}S_2}{(\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2)} \\
&= \frac{2(\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n}))(\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}) - 2\Big(\big(\mathcal{H}_i^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})^2\mathcal{H}_i)\big)\Big)\mathbf{x}S_2}{(\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2)} \\
&= \frac{2\mathcal{H}_i{}^T(\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})\Big[(\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}) - (\mathbf{I} - \tfrac{1}{n}\mathbf{J}_{n\times n})\mathcal{H}_i\mathbf{x}S_2\Big]}{(\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_\mathbf{x}}\right\|_2^2 + \varepsilon_2)}
\end{aligned}
$$

$$(A.34)$$

In terms of the size of each term in A.34,

$$
\mathcal{H}_i{}^T : \mathbf{m} \times \mathbf{n} \quad \mathbf{I} : \mathbf{n} \times \mathbf{n} \quad \mathbf{J} : \mathbf{n} \times \mathbf{n} \longrightarrow \mathcal{H}_i{}^T(\mathbf{I} - \frac{1}{\mathbf{n}}J_{\mathbf{n}\times\mathbf{n}}) : \mathbf{m} \times \mathbf{n}
$$

$$
\mathbf{y}_i : \mathbf{n} \times 1 \quad \mathbf{1}^T : \mathbf{n} \times 1 \quad \overline{\mathbf{y}_i} : 1 \times 1 \longrightarrow (\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}) : \mathbf{n} \times 1
$$

$$
\mathbf{I} : \mathbf{n} \times \mathbf{n} \quad \mathbf{J} : \mathbf{n} \times \mathbf{n} \quad \mathcal{H}_i : \mathbf{n} \times \mathbf{m} \quad \mathbf{x} : \mathbf{m} \times 1 \quad S_2 : 1 \times 1 \longrightarrow (\mathbf{I} - \frac{1}{\mathbf{n}}J_{\mathbf{n}\times\mathbf{n}})\mathcal{H}_i\mathbf{x}S_2 : \mathbf{n} \times 1
$$

$$
\mathcal{H}_i{}^T(\mathbf{I} - \frac{1}{\mathbf{n}}\mathbf{J}_{\mathbf{n}\times\mathbf{n}}) : \mathbf{m} \times \mathbf{n} \quad (\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}) : n \times 1 \quad (\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H}_i\mathbf{x}S_2 : \mathbf{n} \times 1
$$

$$
\longrightarrow \mathcal{H}_i{}^T(\mathbf{I} - \frac{1}{\mathbf{n}}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\Big[(\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}) - (\mathbf{I} - \frac{1}{n}\mathbf{J}_{\mathbf{n}\times\mathbf{n}})\mathcal{H}_i\mathbf{x}S_2\Big] : \mathbf{m} \times 1
$$

$$
\left\|\mathbf{y}_i - \mathbf{1}^T\,\overline{\mathbf{y}_i}\right\|_2^2 + \left\|\mathcal{H}_i\mathbf{x} - \mathbf{1}^T\,\overline{\mathcal{H}_i\mathbf{x}}\right\|_2^2 + \varepsilon_2 : 1 \times 1
$$

$$(A.35)$$

Bibliography

[1] Z. Wang D. Brunet A. Rehman, M. Rostami and E. Vrscay. SSIM-inspired image restoration using sparse representation. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–12, 2012.

[2] Luyi Bai, Changming Xu, and Cong Wang. A review of fusion methods of multispectral image. *Optik*, 126(24):4804–4807, 2015.

[3] Bincy Bavachan and Dr Prem Krishnan. A survey on image fusion techniques. *International Journal of Research in Computer and Communication Technology*, 3(3):48–52, 2014.

[4] Ajay Kumar Boyat and Brijendra Kumar Joshi. A review paper: noise models in digital image processing. *arXiv preprint arXiv:1505.03489*, 2015.

[5] Dominique Brunet. *A Study of the Structural Similarity Image Quality Measure with Applications to Image Processing*. PhD thesis, Waterloo University, 2012.

[6] Virginia Burrus, Charles Burrus, Kalyani Narasimhan, Yongtai Guo, and Caijie Li. Introduction to wavelets and wavelet transforms-a primer, brrus cs, 1998.

[7] R. S. Rajesh C. Morris. Survey of spatial domain image fusion techniques. *International Journal of Advanced Research in Computer Science Engineering and Information Technology*, pages 249–254, 2014.

[8] S.NarayanaReddy C.Chandrasekhar, A.Viswanath. Fpga implementation of image fusion technique using dwt for micro air vehicle applications. *International Journal of Scientific Engineering Research*, 4:307—-315, 2013.

[9] Damon M Chandler and Sheila S Hemami. Vsnr: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE transactions on image processing*, 16(9):2284–2298, 2007.

[10] Nedeljko Cvejic, Artur Łoza, David R. Bull, and Nishan Canagarajah. A similarity metric for assessment of image fusion algorithms. 2008.

[11] O.Michailovich D. Otero, D. L.Torre and E. R. Vrscay. Optimization of structural similarity in mathematical imaging. *Optimization and Engineering*, pages 1–35, February 2020.

[12] Ayush Dogra, Bhawna Goyal, and Sunil Agrawal. From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications. *IEEE access*, 5:16040–16067, 2017.

[13] Geoff Dougherty. *Digital image processing for medical applications*. Cambridge University Press, 2009.

[14] Hammou Dounia, Fezza Sid Ahmed, and Hamidouche Wassim. Egb: Image quality assessment based on ensemble of gradient boosting. pages 541–549, 2021.

[15] Jiao Du, Weisheng Li, Ke Lu, and Bin Xiao. An overview of multi-modal medical image fusion. *Neurocomputing*, 215:3–20, 2016.

[16] Mehran Ebrahimi Kahrizsangi. Inverse problems and self-similarity in imaging. 2008.

[17] Fatma El-Zahraa Ahmed El-Gamal, Mohammed Elmogy, and Ahmed Atwan. Current trends in medical image registration and fusion. *Egyptian Informatics Journal*, 17(1):99–124, 2016.

[18] Anna Geomi George and A. Kethsy Prabavathy. A survey on different approaches used in image quality assessment. *International Journal of Emerging Technology and Advanced Engineering*, 3(2):35–41, Feb 2005.

[19] Subham Ghosh and Yoram Rudy. Application of l1-norm regularization to epicardial potential solution of the inverse electrocardiography problem. *Annals of biomedical engineering*, 37:902–912, 2009.

[20] Maoguo Gong, Xiangming Jiang, and Hao Li. Optimization methods for regularization-based ill-posed problems: a survey and a multi-objective framework. *Frontiers of Computer Science*, 11(3):362–391, 2017.

[21] Maoguo Gong, Hao Li, and Xiangming Jiang. A multi-objective optimization framework for ill-posed inverse problems. *CAAI Transactions on Intelligence Technology*, 1(3):225–240, 2016.

[22] Rafael C Gonzales and Paul Wintz. *Digital image processing*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[23] Christine Guillemot and Olivier Le Meur. Image inpainting: Overview and recent advances. *IEEE signal processing magazine*, 31(1):127–144, 2013.

[24] Ankit Gupta, Neel Joshi, C Lawrence Zitnick, Michael Cohen, and Brian Curless. Single image deblurring using motion density functions. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part I 11*, pages 171–184. Springer, 2010.

[25] Jacques Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*, volume 15. Yale university press, 1923.

[26] Saad Hikmat Haji and Adnan Mohsin Abdulazeez. Comparison of optimization techniques based on gradient descent algorithm: A review. *PalArch's Journal of Archaeology of Egypt / Egyptology*, 18(4):2715–2743, Feb. 2021.

[27] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering [c]. In *European conference on computer vision*, pages 1–14. Springer, 2010.

[28] Alain Horé and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010.

[29] Areeba Ilyas, Muhammad Shahid Farid, Muhammad Hassan Khan, and Marcin Grzegorzek. Exploiting superpixels for multi-focus image fusion. *Entropy*, 23(2):247, 2021.

[30] Jireh Jam, Connah Kendrick, Kevin Walker, Vincent Drouard, Jison Gee-Sern Hsu, and Moi Hoon Yap. A comprehensive review of past and present image inpainting methods. *Computer vision and image understanding*, 203:103147, 2021.

[31] Alex Pappachen James and Belur V Dasarathy. Medical image fusion: A survey of the state of the art. *Information fusion*, 19:4–19, 2014.

[32] Mark D Jasiunas, David A Kearney, John Hopf, and Grant B Wigley. Image fusion for uninhabited airborne vehicles. In *2002 IEEE International Conference on Field-Programmable Technology, 2002.(FPT). Proceedings.*, pages 348–351. IEEE, 2002.

[33] Dong Jiang, Dafang Zhuang, Yaohuan Huang, and Jinying Fu. Survey of multispectral image fusion techniques in remote sensing applications. *Image fusion and its applications*, pages 1–23, 2011.

[34] Harpreet Kaur, Deepika Koundal, and Virender Kadyan. Image fusion techniques: a survey. *Archives of computational methods in Engineering*, 28:4425–4447, 2021.

[35] Zhengfang-HojatollahYeganeh Kede Ma, Duanmu and Zhou Wang. Multi-exposure image fusion by optimizing a structural similarity index. *IEEE Transactions on Computational Imaging*, 4(1):60–72, 2017.

[36] Shiqi Wang Keyan Ding, Kede Ma and Eero P. Simoncelli. Comparison of full-reference image quality models for optimization of image processing systems. volume 4, pages 1573–1405, 2021.

[37] LA Klein. Society of photo-optical instrumentation engineers (spie) 405 fieldston road bellingham. *United States, WA*, 1993.

[38] Shivsubramani Krishnamoorthy and KP Soman. Implementation and comparative study of image fusion algorithms. *International Journal of Computer Applications*, 9(2):25–35, 2010.

[39] Mrityunjay Kumar and Sarat Dass. A total variation-based algorithm for pixel-level image fusion. *IEEE Transactions on Image Processing*, 18(9):2137–2143, 2009.

[40] Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1):011006, 2010.

[41] Long V Le, Young D Kim, and David E Aspnes. Quantitative assessment of linear noise-reduction filters for spectroscopy. *Optics Express*, 28(26):38917–38933, 2020.

[42] Bing Li, Yong Xian, Daqiao Zhang, Juan Su, Xiaoxiang Hu, and Weilin Guo. Multi-sensor image fusion: a survey of the state of the art. *Journal of Computer and Communications*, 9(6):73–108, 2021.

[43] Shutao Li, Xudong Kang, Leyuan Fang, Jianwen Hu, and Haitao Yin. Pixel-level image fusion: A survey of the state of the art. *information Fusion*, 33:100–112, 2017.

[44] Yi Li, Junli Zhao, Zhihan Lv, and Jinhua Li. Medical image fusion method by deep learning. *International Journal of Cognitive Computing in Engineering*, 2:21–29, 2021.

[45] Kui Liu and Guixia Kang. Multiview convolutional neural networks for lung nodule classification. *International Journal of Imaging Systems and Technology*, 27(1):12–22, 2017.

[46] Yu Liu, Xun Chen, Zengfu Wang, Z Jane Wang, Rabab K Ward, and Xuesong Wang. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42:158–173, 2018.

[47] Zhaodong Liu, Yi Chai, Hongpeng Yin, Jiayi Zhou, and Zhiqin Zhu. A novel multi-focus image fusion approach based on image decomposition. *Information Fusion*, 35:102–116, 2017.

[48] Ren C Luo, Chih Chia Chang, and Chun Chi Lai. Multisensor fusion and integration: Theories, applications, and its perspectives. *IEEE Sensors Journal*, 11(12):3122–3138, 2011.

[49] Ren C. Luo and Michael G. Kay. Multisensor integration and fusion for intelligent machines and systems. 1995.

[50] C. De Mol M. Bertero, P. Boccacci. *Introduction to Inverse Problems in Imaging*. CRC Press, 2021.

[51] P.J. Åstrand M. Ehlers, S. Klonus. Quality assessment for multi-sensor multi-date image fusion. 2008.

[52] Jiayi Ma, Yong Ma, and Chang Li. Infrared and visible image fusion methods and applications: A survey. *Information Fusion*, 45:153–178, 2019.

[53] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.

[54] Dhirendra Mishra and Bhakti Palkar. Image fusion techniques: a review. *International Journal of Computer Applications*, 130(9):7–13, 2015.

[55] Jan Modersitzki. *FAIR: flexible algorithms for image registration*. SIAM, 2009.

[56] Pedram Mohammadi, Abbas Ebrahimi-Moghadam, and Shahram Shirani. Subjective and objective quality assessment of image: A survey. *arXiv preprint arXiv:1406.7799*, 2014.

[57] Pinky Tanwar Nupur Singh. Image fusion using improved contourlet transform technique. *Int J Recent Technol Eng (IJRTE)*, 1(2), 2012.

[58] Zaid Omar and Tania Stathaki. Image fusion: An overview. In *2014 5th International Conference on Intelligent Systems, Modelling and Simulation*, pages 306–310, 2014.

[59] Isabel Molina Orihuela and Mehran Ebrahimi. An efficient algorithm for computing the derivative of mean structural similarity index measure. In *International Conference on Image Analysis and Recognition*, pages 55–66. Springer, 2019.

[60] D. Otero, D. La Torre, O. Michailovich, and E. R. Vrscay. Optimization of structural similarity in mathematical imaging. 2020.

[61] Daniel Otero. *A Study of the Structural Similarity Image Quality Measure with Applications to Image Processing*. PhD thesis, Waterloo University, 2015.

[62] Gonzalo Pajares and Jesus Manuel De La Cruz. A wavelet-based image fusion tutorial. *Pattern recognition*, 37(9):1855–1872, 2004.

[63] Nirmala Paramanandham and Kishore Rajendiran. Multi sensor image fusion for surveillance applications using hybrid image fusion algorithm. *Multimedia Tools and Applications*, 77(10):12405–12436, 2018.

[64] Zhong Zhang Rick S. Blum, Zhiyun Xue. *Multi-Sensor Image Fusion and Its Applications (1st ed.)*. CRC Press, 2006.

[65] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[66] Amina Saleem, Azeddine Beghdadi, and Boualem Boashash. Image fusion-based contrast enhancement. *EURASIP Journal on Image and Video Processing*, 2012:1–17, 2012.

[67] Mamta Sharma. A review: image fusion techniques and applications. *Int J Comput Sci Inf Technol*, 7(3):1082–1085, 2016.

[68] Hamid R Sheikh, Alan C Bovik, and Gustavo De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on image processing*, 14(12):2117–2128, 2005.

[69] Le Song, Yuchi Lin, Weichang Feng, and Meirong Zhao. A novel automatic weighted image fusion algorithm. In *2009 International Workshop on Intelligent Systems and Applications*, pages 1–4. IEEE, 2009.

[70] Nahed Tawfik, Heba A. Elnemr, Mahmoud Fakhr, Moawad I. Dessouky, and Fathi E. Abd El-Samie. Survey study of multimodality medical image fusion methods. *Multimedia Tools and Applications*, 80:6369–6396, 2021.

[71] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4:1035–1038, 1963.

[72] Andrei Nikolaevich Tikhonov, AV Goncharsky, Vyacheslav Vasil'evich Stepanov, and Anatoly G Yagola. *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media, 1995.

[73] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of ill-posed problems*. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977. Trans-

lated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

[74] Alexander Toet. Image fusion by a ratio of low-pass pyramid. *Pattern recognition letters*, 9(4):245–253, 1989.

[75] Huaqing Wang, Shi Li, Liuyang Song, Lingli Cui, and Pengxin Wang. An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network. *IEEE Transactions on Instrumentation and measurement*, 69(6):2648–2657, 2019.

[76] Qinxia Wang and Xiaoping Yang. An efficient fusion algorithm combining feature extraction and variational optimization for ct and mr images. *Journal of applied clinical medical physics vol. 21,6 (2020)*, 21(6):139–150, 2020.

[77] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE signal processing letters*, 9(3):81–84, 2002.

[78] Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.

[79] Zhou Wang, Alan C. Bovik, and Ligang Lu. Why is image quality assessment so difficult? In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages IV–3313–IV–3316, 2002.

[80] Zhou Wang, Alan C Bovik, and Hamid R Sheikh. Structural similarity based image quality assessment. In *Digital Video image quality and perceptual coding*, pages 225–242. CRC Press, 2017.

[81] Zhou Wang, Alan C Bovik, and Eero P Simoncelli. Structural approaches to image quality assessment. *Handbook of image and video processing*, 7(18), 2005.

[82] Zhou Wang and Alan C.Bovik. *Modern Image Quality Assessment*. 2006.

[83] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[84] J Yang, Y Ma, W Yao, and WT Lu. A spatial domain and frequency domain integrated approach to fusion multifocus images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(PART B7), 2008.

[85] Chengfang Zhang and Ziliang Feng. Infrared-visible image fusion using accelerated convergent convolutional dictionary learning. *Arabian Journal for Science and Engineering*, 47(8):10295–10306, 2022.

[86] Hao Zhang, Han Xu, Xin Tian, Junjun Jiang, and Jiayi Ma. Image fusion meets deep learning: A survey and perspective. *Information Fusion*, 76:323–336, 2021.

[87] Qiang Zhang, Yi Liu, Rick S Blum, Jungong Han, and Dacheng Tao. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Information Fusion*, 40:57–75, 2018.

[88] H. R. Sheikh Zhou Wang, A. C. Bovik and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.