Depth Map Enhancement through Weighted Guided Image Filters in Shape-From-Focus

Author

Zubair Ahmed

NUST CEME 00000318173

Supervisor

Dr. Ahsan Shahzad

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

JUNE 2022

Depth Map Enhancement through Weighted Guided Image Filters in Shape-From-Focus

Author

Zubair Ahmed

NUST CEME 00000318173

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Computer Engineering

Thesis Supervisor

Dr. Ahsan Shahzad

Thesis Supervisor's Signature: _____

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

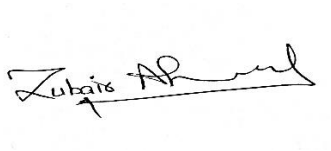
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

JUNE, 2022

Declaration

I certify that this research work titled "*Depth Map Enhancement through Weighted Guided Image Filters in Shape-From-Focus"* is my own work. The work has not been submitted to anybody else for review. The usage of content from other sources has been properly acknowledged and referred to.



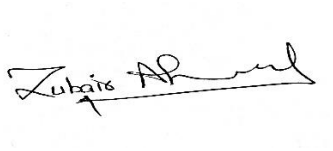
Signature of Student

Zubair Ahmed

NUST CEME 00000318173

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical, and spelling mistakes. The thesis is also according to the format given by the university.



Signature of Student

Zubair Ahmed

NUST CEME 00000318173

Signature of Supervisor

Dr. Ahsan Shahzad

Copyright Statement

- Copyright in the text of this thesis rests with the student author. Copies (by any process) either in full or of extracts, may be made only in accordance with instructions given by the author and lodged in the Library of NUST College of E&ME. Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.
- The ownership of any intellectual property rights which may be described in this thesis is vested in NUST College of E&ME, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the College of E&ME, which will prescribe the terms and conditions of any such agreement.
- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of NUST College of E&ME, Rawalpindi.

Acknowledgments

All praise and glory to Almighty Allah (the most glorified, the highest) who gave me the courage, patience, knowledge, and the ability to carry out this work and to complete it with perseverance and contentment. Undoubtedly, HE eased my way and without HIS blessings I can achieve nothing.

I would like to express my sincere gratitude to my advisor Dr. Ahsan Shahzad and Dr. Muhammad Usman Akram for boosting my morale and for their continual assistance, motivation, dedication, and invaluable guidance in my quest for knowledge. I am blessed to have such a cooperative advisor and kind mentor for my research.

My acknowledgment would be incomplete without thanking the biggest source of my strength, my family. I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout every department of my life and my loving sisters who were with me through thick and thin.

Finally, I would like to express my gratitude to all my friends and the individuals who have encouraged and supported me through this entire period.

Dedicated to my exceptional parents: ***Mr. & Mrs. Bashir Ahmed Sangha****, and adored brothers & sisters whose tremendous support and cooperation led me to this accomplishment.*

Abstract

Typically, shape-from-focus (SFF) approaches do not take into account any prior information in order to improve the depth map's accuracy. Estimation of depth maps delivers a key role in the reconstruction of 3D shapes. There are many monocular approaches that use image focus to reshape 3D shapes, and shape from focus is one of them. It uses information about the focus of the optical system to provide a means of measuring 3D information. This study proposed a framework for the enhancement of the depth map by using a weighted combination set of guided filters in shape-from-focus. It has been observed that a different set of weighted combinations of guided image filters are effective in enhancing depth maps in SFF. After evaluation, it is found that the weighted combination of a set of 2 guided image filters provides an enhanced depth map. In comparison to a recent study in which the authors employed a set of 19 filters to enhance the depth map findings. The proposed study gives better outcomes with a faster and less computations-based framework to boost the depth map. In the literature, many guided image filters have been proposed to enhance the depth map individually, but few of them have computational time flaws, and some have unsatisfactory results. A weighted combination of 2 filter sets has been obtained best filter set combination for enhancement of the depth map after evaluation. The optimized weights are obtained using the particle swarm optimization approach, and the subset of best-performing filters is identified through a sequential forward search method. The experimental results have demonstrated that the proposed framework provides considerably improved depth maps, yielding 93% correlation and 4.7 root mean square error to the actual depth map.

Keywords: *Depth map estimation, Guided image filtering optimization, Shape-from-focus, Focus measure, Particle swarm optimization, Ground truth.*

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

As automation and artificial intelligence become increasingly popular, so does the need for accurate computer vision and scene rendering. Depth estimate is one of the most significant aspects of computer vision applications. One of them is determining the 3D shape of an item from its 2D visual input. Focus analysis has been an important tool for retrieving the 3D structure of objects, with various methods for retrieving 3D shapes. Schemes based on focus analysis are further classified into two types: shape-from-focus (SFF) and shape from defocus (SDF). SFD methods use less numbers of images which are taken with multiple focus perspectives. The relative blur among these images used by the algorithm to estimate the 3D structure of the object. While, SFF techniques create a compact depth map from a series of images captured at various focus levels. Shape-from-focus has been extensively researched in computer vision as a method for affected depth recovery and 3D reconstruction, and it has found success in a variety of industrial applications. Acquiring depth estimation has been identified as a major topic in the field, and numerous ways to extracting the scene's depth have been offered. However, when the acquired images lack both color and feature information depth estimate using stereo matching becomes difficult and unreliable. To recover improved or enhanced depth maps, the proposed method uses guided filtering in SFF. The depth enhancement procedure can be guided by the image sequence and image focus volume. Guided image filtering is a method for modifying or enhancing an image using weights based on the attributes of another image known as a guidance map. This technique filters the input image using structural information from the guidance image. The information in both the input image and the guide map images is about the same scene, but it is in different domains.

1.1 Motivation

There is some work on depth estimation by a variety of methods, for example, color filtered dual camera, single-camera video clip processing, and dual camera on a mobile phone. Our research is motivated by the great benefit of guided filtering, where our goal is to improve the depth map of images using guided filtering in SFF that will help to reconstruct the 3D shape of the object. The performance of focus measures is influenced by several aspects in the images including noise, window size, imaging agent properties, and texture. The object's detailed structures may be missing from the reconstructed depth or 3D shape.

In the shape-from-focus literature, many methods have been proposed to undertake these problems, and these methods are distributed into two sections. In the primary techniques for enhancing the preliminary focus measure image and presenting an improved image of the focus measure. These approaches then use the straightforward “Winner Takes All” methodology to estimate the depth image. In the secondary techniques, make a preliminary depth image from the image focus measure, and after that refine it to create an improved depth image/map.

1.2 Problem Statement

In computer vision and computer graphics, 3D shape reconstruction is a technique for capturing the form and appearance of real-world objects. There are several uses for 3D object reconstruction, from engineering and medical to creative and commercial design. Because it's being utilized in a broad variety of image processing and computer vision applications. An object reconstruction will be more accurate with a higher detailed photograph as the application's primary pillar. An object reconstruction approach relies heavily on a depth map. For example, light, window size, noise, scene texture and picture attributes all have an impact on depth map quality. As a result of these issues, the focus measure generally provides erroneous information about the focus of attention. As a consequence, the reconstructed depth or 3D shape may be devoid of the particular structures of the item or scene. To address these issues, we decided to focus on this area for our study, and to create a better depth map of the objects or scenes that would aid in their reconstruction.

1.3 Aims and Objectives

The main purpose of the proposed research is to review and study fresh developments in depth map enhancement and improve the depth map of images using guided filtering in SFF, by incorporating simple methodology with fewer computations. Guided filters are used to estimate the depth of synthetic objects using image processing filtering techniques. Make an image depth map enough to make the better decision or make him able to estimate the distance between the subject and the object.

1.4 Area of Application

Many computer vision applications require the acquisition of depth information about a scene. In Robot Vision, Human-Computer Interface, and many other vision-based

applications, the depth map plays a vital role. The performance of 3D image acquisition and Intelligent Driver Assistance Systems is also influenced by the scene's depth map. Action recognition, autonomous navigation, virtual-augmented reality and 3D television are all areas in which they play a significant role.

1.5 Structure of Thesis

This work is structured as follows:

Chapter 2 Briefly represents the image depth map and how the depth map is conventionally estimated. In addition, an overview of guided image filters and their various categories.

Chapter 3 Covers the analysis of the existing research work done for the estimation of depth map in past years.

Chapter 4 Describes the proposed research work in detail. This section also defines the method undertaken to enhance the depth map using guided image filters in shape-from-focus.

Chapter 5 Provides the review of datasets and the performance measures used for interpretation of the proposed research work. All of the results are well-explained, along with the required tables and figures.

Chapter 6 Concludes the framework and introduces the future scope of this research work.

CHAPTER 2: DEPTH MAP AND GUIDED IMAGE FILTERS

Computers are becoming more and more like spike, causing machines to a more debatable and understandable height. Artificial intelligence has been a part of our world for a few years now. It is inconceivable to disregard the ability of machine learning. Proper use of artificial intelligence can change the world, by making it simple to process large amounts of data and make meaningful judgments. Many alternative methodologies have been proposed in order to discover new and more precise methods of determining the distance between objects, whether it is to show the way, to avoid obstacles, or to turn.

2.1 How to estimate depth map?

The traditional method of determining the distance of an item in an image needs a rectified stereo epi-polar lines image like the normal human eye. Which can be used to extract geometric information from a scene using parallel line, called parallax. The distance to an object can be calculated which is commonly called the depth of the scene. With the help of different images of the same scenes from a different point of view. The distance between these points and the camera can be calculated by calculating the displacement of points across different images. The effects of parallax can be seen in the differences of the two images points in the stereo pair. When two horizontally displaced cameras image planes are projected onto a point in the scene. The nearer the point is to the baseline of the camera, the greater the difference in its relative position on the image planes. The purpose of stereo matching is to find the relevant points and calculate their displacement. In order to reproduce the scene's geometry as a depth map.

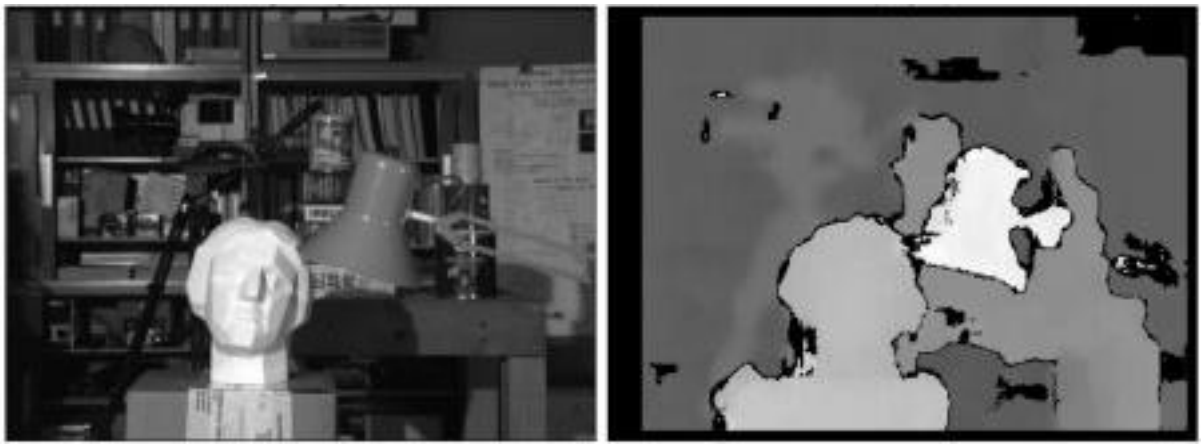


Figure: 2.1 Depth map of a scene

2.2 Depth Map

The distance is represented by an image or image channel between the surfaces of objects from a point of view is called depth map. Figure 2.1 shows the initial depth of the scene where the intensity of the object closest to the subject is brighter while the intensity of the objects as have dark intensities. Calculating depth is an essential feature of understanding geometric relationships within a scene. Figure 2.2 shows the depth map of the scene on the left, in which the original color channel of the scene can be seen. While the initial depth map on the right side of the image estimates that this is the disparity output. With this kind of depth map machine cannot make a decision nor can it estimate the distance between the subject and the object.

Figure: 2.2 Original scene vs disparity depth map of scene

For many computer vision applications, obtaining depth information about a scene is fundamental such as navigator, Autonomous Navigation, Robot Vision, virtual-augmented reality, Human-Computer Interfaces, Intelligent Visual Surveillance, Intelligent Driver Assistance Systems, 3D Image Acquisition, Synthetic Focus, 3DTV and Action Recognition.

2.3 Guided Image Filters

In computer vision and computer graphics, 3D shape reconstruction is a technique for capturing the form and appearance of real-world objects. There are several uses for 3D object reconstruction, from engineering and medical to creative and commercial design. Because it's

being utilized in a broad variety of image processing and computer vision applications. An object reconstruction will be more accurate with a higher detailed photograph as the application's primary pillar. An object reconstruction approach relies heavily on a depth map. For example, light, window size, noise, scene texture and picture attributes all have an impact on depth map quality. As a result of these issues, the focus measure generally provides erroneous information about the focus of attention. As a consequence, the reconstructed depth or 3D shape may be devoid of the particular structures of the item or scene. To address these issues, we decided to focus on this area for our study, and to create a better depth map of the objects or scenes that would aid in their reconstruction. Some of them are texture removal [19], [20], extraction of joint structure [18], up sampling the depth image, and reduction of haze [17]. If there are any additional details related to the scene that is present in the guidance image/map but not in the input image called the target image. These additional details can be redirected from the reference image to the target image. This is the mainly key benefit of the guided filters.

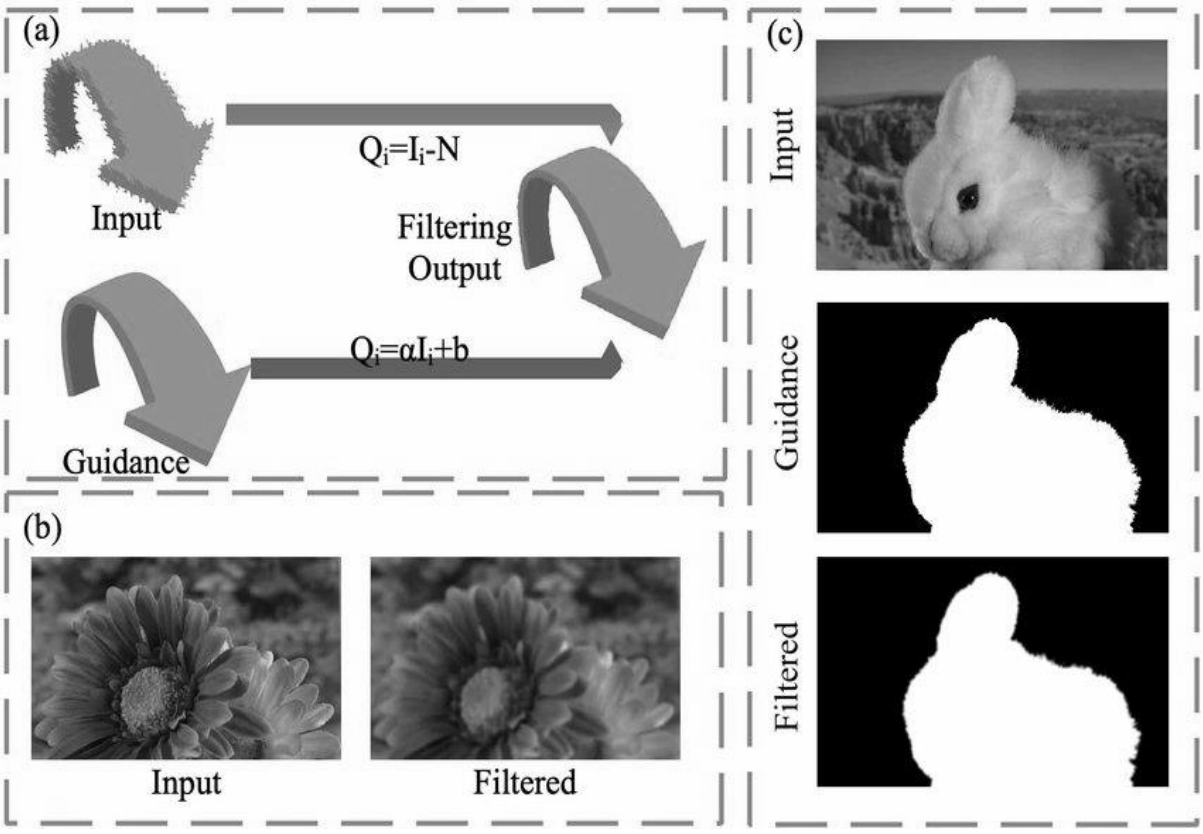


Figure: 2.3 Guidance image and filtered output

In Figure 2.3 both images guidance and target images (input image) can be seen. This extra info may be useful in enhancing the target image's filtering. Two assertions are used to validate the information transfer: one is the guide image contains flawless structural information that matches the target image. In second assertion, guidance map can give more trustworthy info, such as edges information.

The issue occurs, however, when the target and guide images are structurally incompatible. This happens when different sensors are used to acquire both the image input and guidance (RGB color issue [18]). Also, in various lighting circumstances with daylight or night [18] and with sensor flash or without sensor flash [21]. It is possible to correct this mismatch by adding a restriction that prohibits unneeded or contradictory data from being exchanged between the guide picture and the target image. It's now up to us to find a way of transporting valuable information without distorting it to the point of becoming uncomfortable.

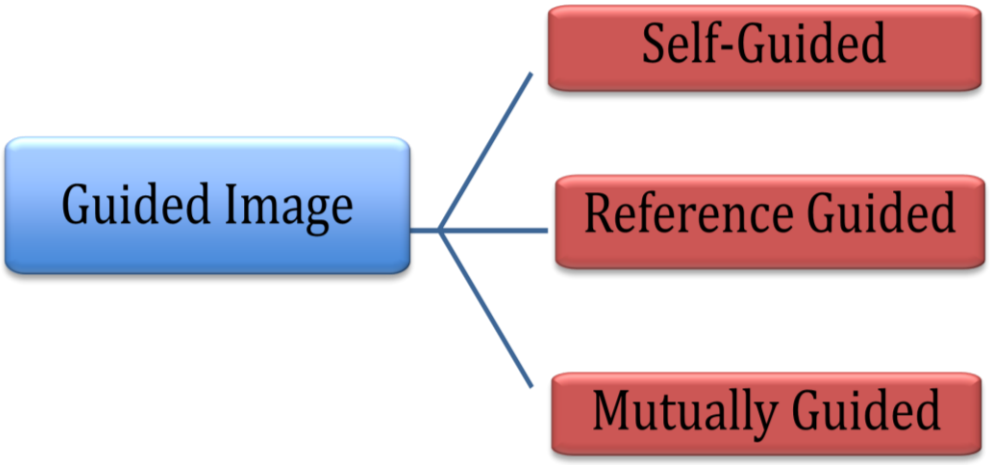


Figure: 2.4 Guided image filter types

We measured the effectiveness of 19 guided image filters in this study. These filters are part of a large family of approaches for guided filtering. As shown in Figure 2.4, we divide guided filters into many types. We split guided filters at the first level of categorization based on the number of images on which they perform the filtering process. Those images are integrated to modify the input images. Self-guided, use only a single image in the filtering process. Whereas reference-guided filters, and mutual guided fall into the multiple images category in which more than one images are engaged in the filtering process. Multiple images-based filters use more than one image in filtering process.

2.3.1 Self-Guided Filters

As the name implies, the target picture (input image) is modified by calculating filtering weights from itself in this category. Whether the guiding picture keeps varying or stays static. In BF (Bilateral Filter) [22] weights stay static and vary in rolling guidance image filer [23]. Self-guided filters are additionally divided into two categories static-self and dynamic-self guidance systems. Inherently iterative, the dynamic approach posits that the filtered input (possible output) provides more precise weights than the initial input. As a result, the filtering weights for each iteration in this method are created from the filtered output of the previous iteration.

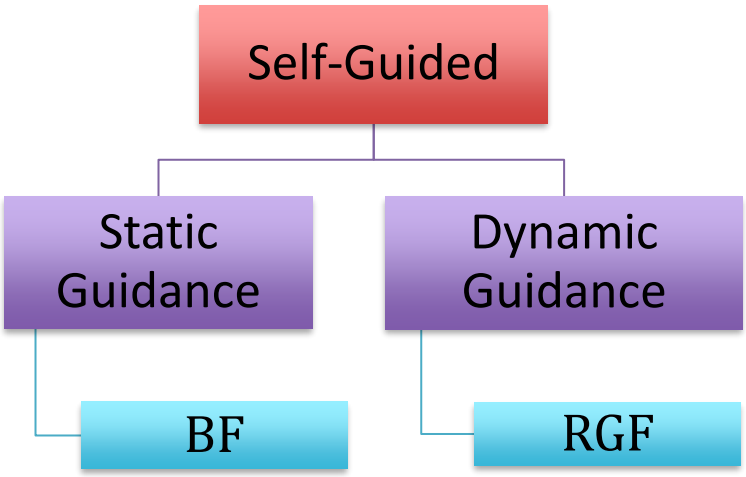


Figure: 2.5 Self-guided filter types

2.3.2 Reference Guided Filters

In the reference based guided filters, a reference image is used to filter the input image. Static and static/dynamic guidance approaches have been classified into this group. The reference in both strategies remains constant. In the static-dynamic guidance approach, the filtered target image is used recursively to create multiple weights from the reference image in addition to compute static weights. As a result, this method uses static guidance to combine the attributes of the reference image and dynamic guidance to combine the properties of the filtered target. The static-reference guided category encompasses the majority of the guided filters previously proposed. This group is further divided into global and local methods, depending on the techniques that is used, whether they use local filtering or global optimization techniques. The majority of local approaches are joint extensions of

single image smoothing and edge-preserving filters [18].

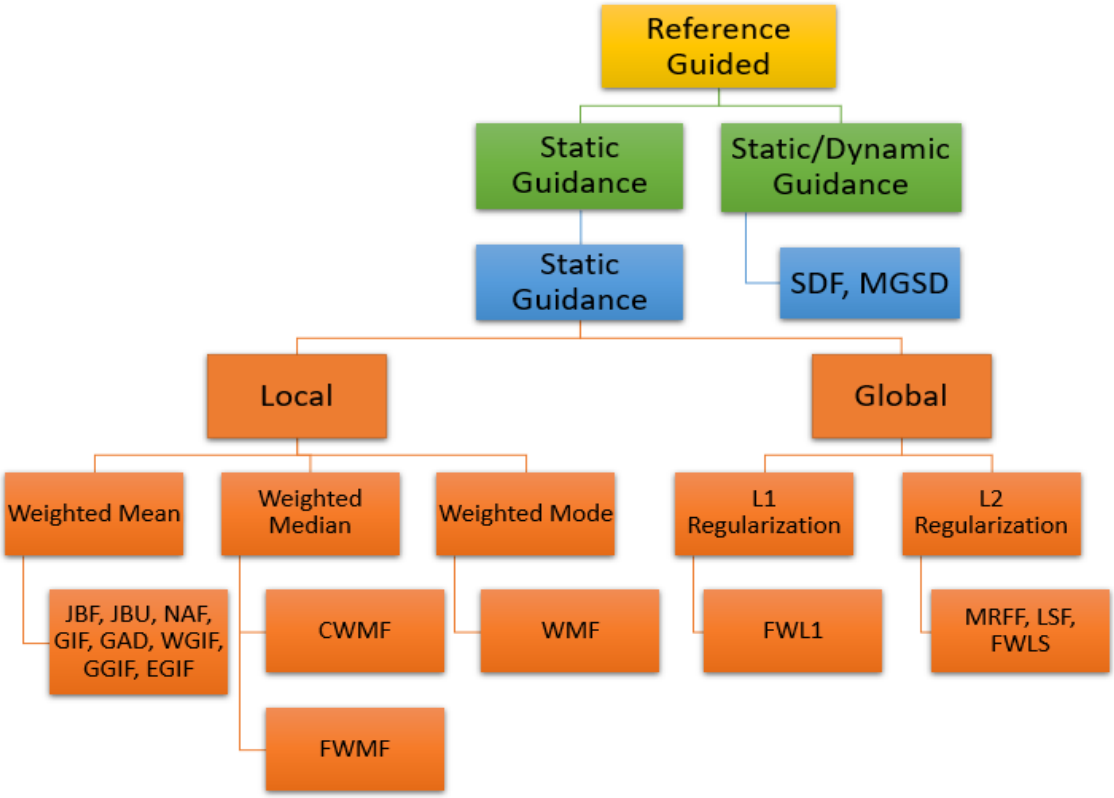


Figure: 2.6 Reference guided filter types

The weights used to filter the input target image are based on the similarity of pixels in the reference image within a local window. Median filtering, weighted mean filtering, and mode filtering procedures are among the local strategies that can be used. These weighted mean approaches can be stated as follows in a generalized linear translation variant filtering process including an input image (target) *T*, an output $\hat{\mathrm{T}}$ and a guidance image *G*:

$$\hat{\mathrm{T}}_p = \sum_{\mathrm{q}\varepsilon\mathrm{N(p)}} W_{pq}(G)T_q \qquad (1)$$

where $W_{pq}(G)$ are weights calculated from the guidance image *G* and $\sum_{\mathrm{q}\varepsilon\mathrm{N(p)}}$ are pixel locations. Non-average filters can also be used to accomplish edge preservation filtering. The pixel value is replaced by the weighted median and weighted mode of its neighbors in weighted median or mode filters, which are generated from the reference image. In terms of computation, local filtering algorithms are typically efficient. Although local filters might cause artefacts at edges (e.g., halo, gradient reversal [17]). On the other hand, various methodologies create filtering in the global optimization context. A reliability

expression and a prior regularization (smoothness) expression are typically included in the optimal solution of these approaches. After calculating the regularisation weights from the guidance image, the output is computed similar to local techniques. These techniques can be summarized as follows:

$$min_{\hat{T}} (\hat{T}, \mathrm{T}) + \lambda\phi (\hat{T}, \mathrm{G}) \quad (2)$$

where input image (target) *T*, an output $\hat{\mathrm{T}}$, a guidance image *G*, and λ represent positive coefficient that governing the relative status of reliability and prior regularisation. expression. The first part of the equation 2 shows data fidelity and the second part denotes regularization. The key difference between global techniques is the regularisation. term, which is commonly defined. in the weighted. L1 or L2 normalization [18]. The L2 normalization is used by the majority of approaches in the literature due to its speed of computation. In most cases, global techniques effectively eliminate edge blurring, but at the cost of global intensity changing [17].

a) Local Reference Guided Filters

Adding pixels in intensity or range and space or domain according to the Gaussian distance is done by the bilateral image smoothing filer described in [22]. High-contrast edges are preserved when using this filter or one of its variations. Low-contrast edges or gradual transitions are removed. Distances from the guiding picture are calculated via the Joint Bilateral Filter (JBF) [21]. The Joint Bilateral Up sampling Filter (JBU) presents a comprehensive architecture for multi-model picture improvement. A high-level-resolution color photographs are utilized as a starting point to create a superior high-resolution depth map. With the Noise Aware Filter, you may boost the quality of depth maps by adapting to noise [26]. When the Weighted Mode Filter (WMF) [27] is used to create a combined depth map histogram, the quality of the depth map is improved. Guided Image Filter (GIF) [17] is one of the quickest edge-preserving smoothing filters. The GIF filter assumes that the output picture is a local linear transformation of the guiding image during the filtering process. The discrepancy between the input image's linear transformation and the guidance provided by a GIF filter is reduced. Optimizing utilizing the fidelity expression lowers the discrepancy. JBF gradient reversal artifacts are also prevented by the GIF filter [21]. Halos may be a problem because of the fixed regularization value. Guided Anisotropic Diffusion (GAD) [28], based on a heat diffusion mechanism, is the definition of depth augmentation. Color image pixels

are used to control the diffusivity of known depth data in this head diffusion system. Instead of weighted guided image filter, gradient domain GIF (Gradient-domain Guided Image Filter) [32] incorporates a first-order edge aware restriction to avoid halo artifacts. Outliers like salt and pepper noise may be eliminated by using the median image filter. The median value of a pixel's neighbors is used to replace the pixel's original value in a median image filter. There are two GIF upgrades suggested by the Effective-GIF (Effective Guided Image Filter) [33].

As a first step, it incorporates an average of the local fluctuations of each pixel into the guided image filter's cost function. Secondly, a content adaptive magnification factor for the detail layer is computed to reduce noise while magnifying tiny details in the layer. Color and depth information should be merged to generate low noise and high-resolution maps of depth, according to the creators of Markov Random Fields-based Filter (MRFF) [34]. Edge changes and depth interruptions tend to align; therefore**,** integration is based on this idea. The Markov random field formulation's minimization problem is solved using a conjugate gradient technique. Large Sparse Fusion (LSF) [35] is a technique for combining sparse laser data with an image to produce a dense depth map. The only other smoothing term anticipated was in [34], and that was just for the first order.

b) Global Reference Guided Filters

Least Squares (FWLS) Fast Weighted L1 (FWL1) filters [24] divide the optimisation issue into a series of one-dimension sub-difficulties for global edge-preserving smoothing. This simplification allows for the use of efficient WL1 and WLS smoothing algorithms. According on the evaluation of similarities between target and guidance. Mutual Structure Joint Filter (MSJF) [18] refreshes both the target input and guidance image. By combining the benefits of both static and dynamic guiding, the Static Dynamic Filter (SDF) [20] provides a unified filtering system. The input target image modified based on a weight function that is solely dependent on the guidance image in static guidance filtering. The guidance image remains fixed in static filtering. Dynamic filtering [23] iteratively constrains the output to reduce structural differences by using the filtered target picture. The findings are provided by Dynamic-Dynamic Guided Filter (MDDG) and Mutual Static-Dynamic Guided Filter (MSDG) [19], which leverage the guidance in a similar fashion to MSJF and SDF. MGDD, on the other hand, prevents halo artefacts (which are common in MSJF due to the formulation of local filtering).

Table 2.1: List of 19 guided filters that are used in this study.

No.	Filters	Abbreviations
1	MSJF [18]	Mutual Structure Joint Filter
2	MRFF [34]	Markov Random Field-based Filter [5]
3	FWL1 [24]	Fast Weighted L1
4	LSF [35]	Large Sparse Fusion
5	WGIF [31]	Weighted Guided Image Filter
6	FWMF [30]	Fast Weighted Median Filter
7	GAD [28]	Guided Anisotropic Diffusion
8	GGIF [32]	Gradient-domain Guided Image Filter
9	EGIF [33]	Effective Guided Image Filter
10	CWMF [29]	Constant-time Weighted Median Filter
11	FWLS [24]	Fast Weighted Least Squares
12	MSDG [19]	Mutual Static-Dynamic Guided Filter
13	SDF [20]	Static Dynamic Filter
14	MDDG [19]	Mutual Dynamic-Dynamic Guided Filter
15	JBU [25]	Joint Bilateral Up-sampling
16	NAF [26]	Noise Aware Filter
17	WMF [27]	Weighted Mode Filter
18	GIF [17]	Guided Image Filter
19	JBF [21]	Joint Bilateral Filer [24]

We assessed the implementation of all the guided filters explained above in this work, with the exception of BF [22] and RGF [23], which do not use guidance image. It's worth noting that some of the guided image filters used in this study were designed for depth upsampling in the first place. As a result, they've been tweaked and tailored to function nicely with our SFF issue. Even though their upsampling factor was fit to one, they were able to use their guided depth-denoising capacity.

2.3.3 Mutual Guided Filter

During the filtering process, this collection of filters changes both the target picture and the reference image. During the filtering process, the target picture serves as a point of reference for the reference image. Filtering weights are computed from the target and reference images iteratively in the dynamic-dynamic approach. After computing weights, combine and update both photos. These filters transmit structural information between target and guiding pictures based on structural consistency (i.e., sharp edges and smooth regions).

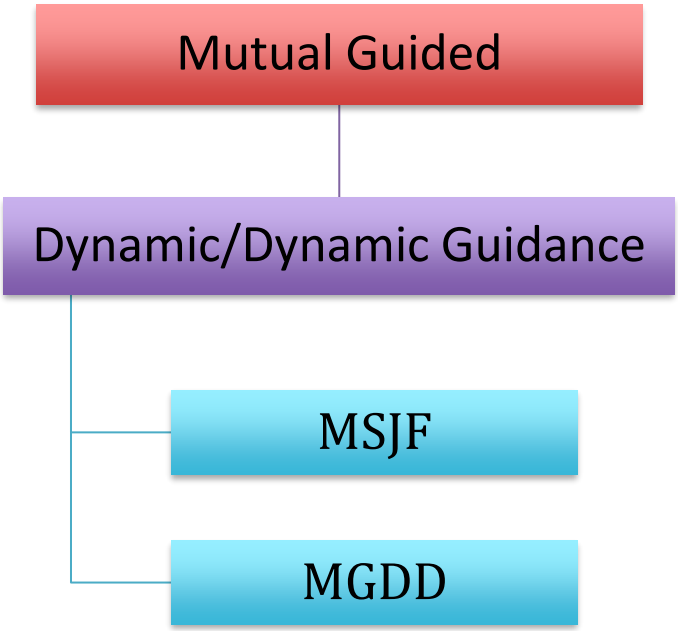


Figure: 2.7 Mutual guided filter types

The results are provided by Mutual Static-Dynamic Guided Filter (MSDG) and Dynamic-Dynamic (MDDG) filters [19]. They use the guidance in a similar fashion to MSJF and SDF. Mutual static dynamic guided filters prevents halo artefacts (appear in MSJF due to local filtering formulation).

CHAPTER 3: LITERATURE REVIEW

Shape-from-focus (SFF) techniques can be arranged in several steps as represented in Figure 3.1. Initially, several images of the scene are captured with a different focus setting of the imaging gadget. These images with different focus measures are along z-direction stacked. The subsequent arrangement of images is typically known as an image sequence or focal stack. As per the law of lens, the best focused pixels captured by the image sensor provides the best depth data for those pixels. Thus, in the second step of shape-from-focus the focus characteristic of every pixel in the image stack is estimated by using the focus measure (FM) operator which is a predefined operator used to apply on each pixel in the image stack. The output of the FM operator is known as initial focus volume. There are many factors which can affect the implementation of the focus measure operators like noise level, texture, illumination, imaging gadgets, or contrast [5]. The scene's accurate structures or information may be missing from the 3D shape or depth of the scene. Because of these aspects initial focus volume contains imprecise focus values. And the reconstructed depth information or 3D shape could come up short on exact constructions of the object from the scene. To handle these issues, various strategies have been suggested in the SFF literature.

In the literature, numerous optical studies have been proposed for the estimation of depth map. These studies mainly categorized into two forms. In the first category, initially enhance the initial focus volume and from initial volume image estimate the depth information by using depth enhancement strategy. Whereas in the second category, initial depth map of the scene is calculated from image focus volume and then process the initial depth map to obtain improved depth image. These processing steps have been shown in Figure 3.1.

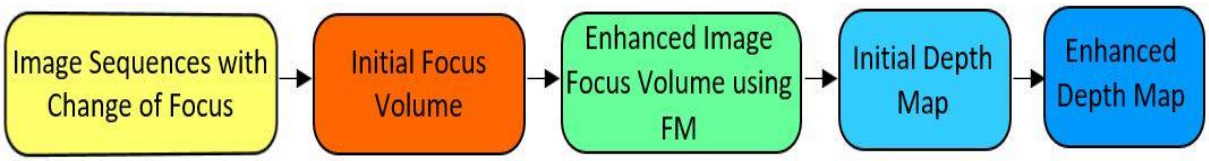


Figure: 3.1 Shape-from-focus System Steps

A fundamental problem with all of these methods is that if the initial depth map estimates are largely incorrect, some enhancement can be expected from these techniques, especially when no additional information is considered.

This study proposes the implementation of guided image filtering in SFF to retrieve better depth maps. The guidance information could be used from the image layout or image focus volume. This guidance image guides the process of increasing the depth map. Basically, this study encourages the successful use of guided image filtering, for which several effective methods have recently been suggested in the previous studies. Guided image filtering is a filtering process that improves an image known as target image. By using certain weights that depend on the properties of another image that is commonly referred to as a guidance image or a guidance map. Both images guidance map and target images have similar scene properties, but their information can be in different dimensions.

3.1 Existing Techniques

Min *et al.* [6], introduces a new way to increase video depth. Provided a high-resolution color video and related low quality in-depth video. The proposed based on combined histogram which is used to improving the quality of depth by suppressing noise and increasing its resolution. The weight calculation is based on the color match between the reference on the color image and the neighboring pixels and is then used to count each one on the combined histogram of the depth map. While the ultimate solution on the histogram is calculate by Global Mode Search. They also reveal that temporarily increasing the depth of video solves a glittering problem. That improves the accuracy of depth video.

Gong *et al.* [7], for color images inadequate depth maps are available they perform sampling and inpainting. They create a heat dissipation-based problem. In which pixels with known depth values are considered as sources of heat and depth is increased by spreading depths in unknown regions. They added the problem of the steady-state of this spread to the popular random hike model, which effectively improves by solving a sparse linear system.

Sun *et al.* [8], they show that weighted median filtering achieves comparative accuracy with various sophisticated accumulation methods for correcting differences. This is due to the well-weighted median filtering features to eliminate outlier errors while respecting edges/structures. The previously overlooked refinement can be at least as important as the

accumulation. They also developed the first permanent time algorithm for a weighted median filter that uses more time than ever before. This simple combination of "box aggregation + weighted median" makes it a practically attractive solution for both speeds and accuracy. They show their dominance in different applications like depth clip-art JPEG artifact removal, up-sampling, and image stylization.

Jacob *et al.* [25], multiple images of a focused object are compared to get in-depth information from the camera's IR emitter, which creates a recognized IR pattern. The BMSV RGB-D camera's depth map and the generated depth map are linked. They propose a method for generating a functional magnetic resonance imaging (fMRI) using 3D realistic head models. The results revealed an increase in the RGB-D sensor's speed and accuracy, resulting in a dense density of correct depth information regardless of the object's surface.

Domanski *et al*. [26], they propose a new parallel technique that significantly reduces the time it takes to estimate depth maps. In this method, multiple arbitrary position input views are used to simultaneously create depth maps. Estimates are made for the parts and their sizes are used to control the trade between the quality of the depth maps and the processing time of the depth estimates. They also suggest a way to improve the temporary consistency of depth maps. This technique uses time-predicted depth, thus estimating the depth for P-type depth frames. For such depth frames, the temporal consistency is higher, while the complexity of the estimation is relatively less.

Aizawa *et al*. [27], they propose a novel alpha channel estimation algorithm for seamless blending with the method of improving the depth map for hairy scenes. Existing matting algorithms can be significantly improved by using additional depth or infrared (IR) information. They further developed the method of alpha estimation in the temporary domain. The depth map was augmented by filtering spatiotemporal locations as well as depth values based on information provided by color and alpha images. The proposed method was tested primarily using the Time of Flight (TOF) camera, and Kinect is also used. Experimental results showed that the proposed method could produce a 3-D scene with a higher degree of naturalness than other methods.

Shang *et al.* [28], they proposed a novel depth-guided affine transformation be used to filter the properties of irrelevant intensity, which is used to improve the properties of the depth map. Since the quality of the initial depth features is low, in-depth guidance filtering of

intensity properties and refining of intensity-guided depth features is done repetitively, which is done gradually and promotes the effects of such actions. In the refinement, depending on the frequency of the depth characteristics, repeat the above Sub-network, as well as outstanding learning, has been introduced.

Usman *et al.* [37], they propose to improve focus measure by developing an energy minimization framework that integrates two types of shape priors and uses a nonconvex regularizer. The proposed regularizer is resistant to focus values that are noisy. The input picture sequence shape prior is the first one proposed, and it is a single and static shape prior. A sequence of form priors makes up the second shape prior. FVs are used to repeatedly create these shape priors. Both of these shape priors confine the solution space for the output focus measure. To optimize no convex energy functions, we employ a memorize-minimization strategy, which ensures a local minimum repeatedly and quickly converges.

Guided filtering is important because it can transmit information from the guidance image to a target image if it has information that isn't present in the target image. When there are structural variations between the target and guide pictures, the issue arises. Typically, this occurs when both images input image and guidance image are captured by different devices in various lighting circumstances with daylight or night [18] and with sensor flash or without sensor flash [21].

To resolve this discrepancy, few barriers can be implemented to prevent the shifting of irrelevant or conflicting information from the guidance map to the target image. Consequently, the difficulty is how to use helpful information wisely without introducing intolerable distortions. Looking at the recently proposed images filtering technique, it has been observed that these studies ignore shape-from-focus in their applications, except for one recently proposed study. In which Usman et al. used 19 different guided image filters to obtain an enhanced depth map of the scene [36]. While they are achieving 91% correlation and 5.05 root means square error with respect to the original depth of the scene by using average results of all these 19 filters. This technique has required more computations and a single guided filter may not be sufficient to enhance the depth map. In this research, we present shape-from-focus as an implementation with guided image filters for the enhancement of depth map. A strong guidance map can be created by obtaining information from image focus volume and a guidance map performs an essential part in guided filters. Though, the choice of guidance map for depth improvement in SFF is not simple. Therefore,

we use the mean image intensity as a guidance map. It has been observed that from thirteen different guidance map mean image intensity guidance map perform better and transfer more information as compared to others [36].

Above all, a brief relative study of common and recently proposed studies based on their ability to improve depth maps is described. In this SFF framework, we suggest a relative analysis of guided image filters for enhancing depth. The implementation of guided image filters has been categorized by examining different methods.

This research shows that guided image filtering is valuable for enhancing depth map. Guidance map based on average image intensity are determined with optical dimensions and 19 guided filters, as a recent study suggests that good performance is provided regardless of the underlying complexity of the scene or noise level [36]. From these 19 guided filters, this study provides a weighted set of filters that performs better for enhancing depth map than all other sequential combinations of 19 filters. These best filter combinations are obtained using Sequential Forward Search (SFS) approach [39] and the optimized weights for these combinations are obtained from Particle Swarm Optimization approach [40]. For further details read relative section in methodology chapter.

CHAPTER 4: METHODOLOGY

In this section, we will go over the proposed framework for improving depth maps as well as detailed steps involved in the framework. The overlook of the proposed framework is depicted in Figure 4.1. Initially, an FM operator is used to obtain a focus volume for an input image sequence. Then, from the output of the focus measure operator, an initial depth map is projected. The mean image intensity in the image sequence is a better choice as guidance map adopted from [36]. It is made up of image sequence and focus volume. Afterward, guided image filters are fed with an initial depth map and an appropriate guide map as input. This study proposes a set of 2 best filters for each synthetic image, taking the output of these 2 filters that are chosen by using sequential forward search SFS approach to selecting the best performing filter from 19 filters. These 19 filters are proposed as a set of guided filters for the enhancement of depth map [36]. Feed them into particle swarm optimization (PSO) algorithm, where optimized weights for each filter are obtained. At the end, multiply these filtered outputs with optimized weights that are obtained from the PSO algorithm and sum up all the results. Finally, the resultant depth map is the enhanced depth map as compared to the initial depth map.

4.1 Initial Focus Volume and Depth Map

Initial depth map for given input images, focus or level of sharpness for each pixel is estimated using an FM operator. Accordingly, an image initial focus volume F*i* (x, y) is given by:

$$F_i(u, v) = FM \otimes I_i(u, v) \tag{3}$$

where *FM* denotes to a focus measure operator, $\otimes$ is the 2-D convolution operator, *i* represents image sequence number and $I_i\ (u, v)$ is the position of the pixels from the image. A notable focus measure operator is GLV [3] that measure the fluctuation of image intensities inside a little nearby window, and is given by,

$$F_i^{(GLV)}(u, v) = \frac{1}{|N|}\sum_{(u,v)\in N}(I_i(u, v) - \mu)^2 \tag{4}$$

where μ and $|N|$ is the total number of pixels and mean gray-level respectively in the local window *N* centered at *(u, v).* This proposed framework can be used any FM operator.

Furthermore, linear filtering is applied to this initial focus volume to control the noise concentration measurement. For this 5 × 5 neighborhood focus values are aggregated, and it results in an enhanced focus volume.

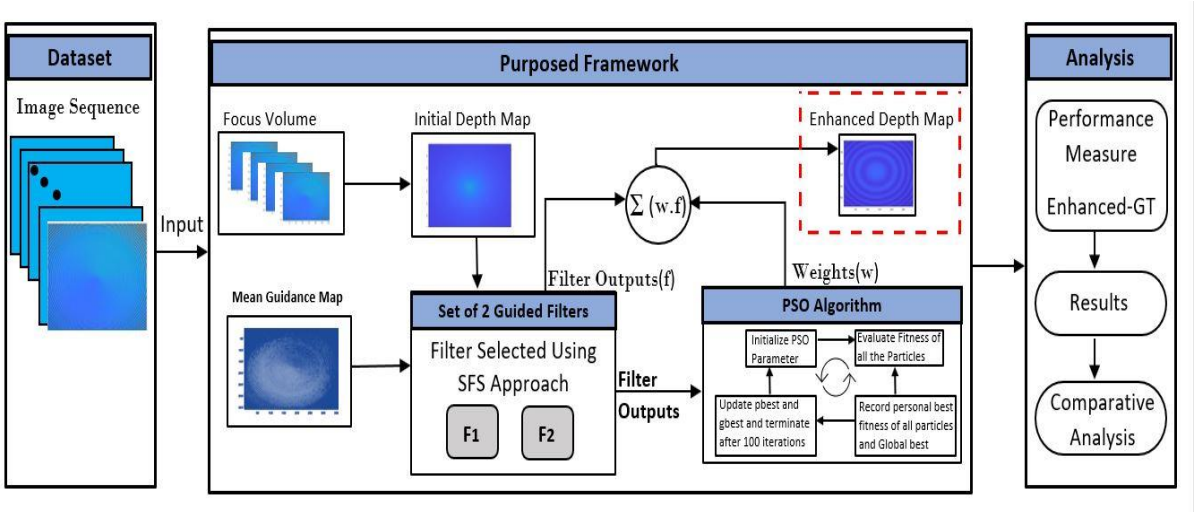


Figure: 4.1 Proposed framework for depth map enhancement through Weighted Combination of Guided Filters

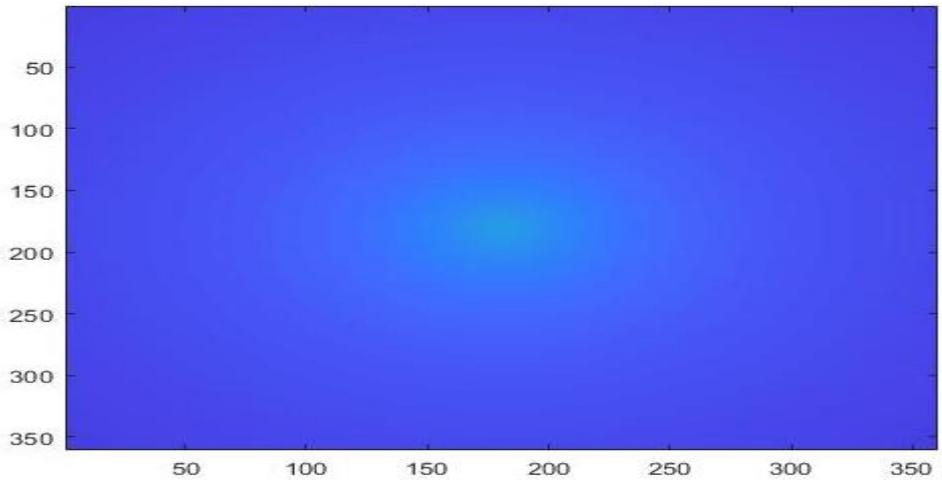
With better focus measurement, the depth map is created by locating the image number according to the highest focus measurement for each pixel location on image *I(u, v)* [1]. The resulting yield is called the initial depth map, shown in the Figure 4.2(a), (b), and (c) for cone, sine and cosine images and given by,

$$D_{(initial)}(u,v) = arg_i max\left(F_i^{(FM)}(u,v)\right) \quad (5)$$

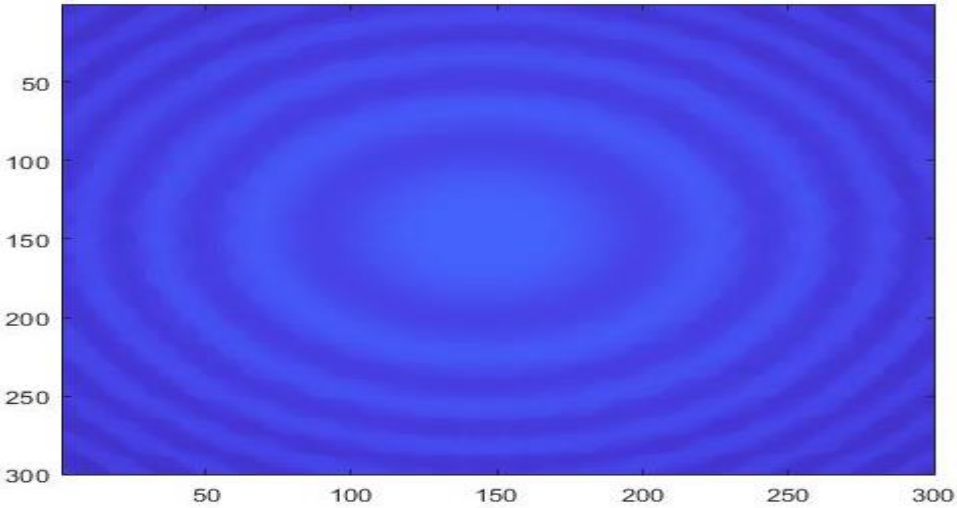
where $F_i^{(FM)}$ is the focus measurement as described in the equation 1 applied to each pixel *(u, v)* of the image. And $arg_i max$ is the maximum value for the corresponding window of focus measurement when applied to image *(u, v)*.

4.2 Guidance Map

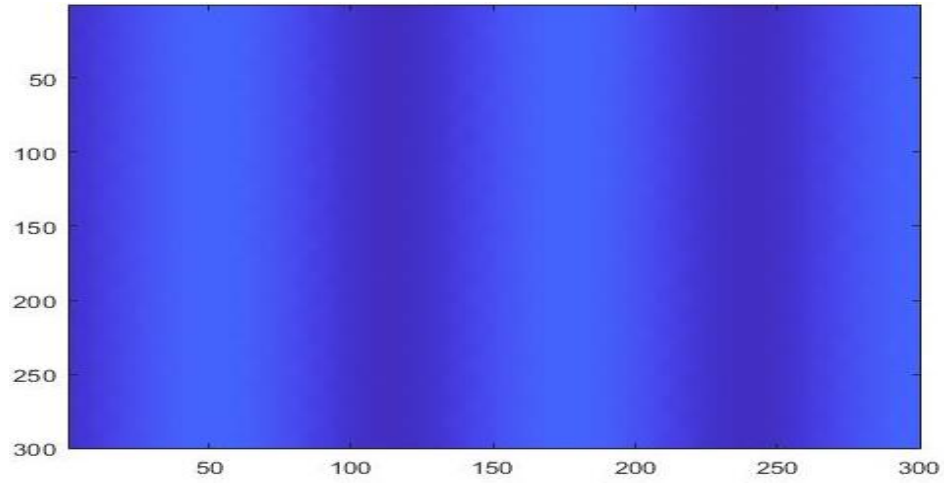
Due to the limited capabilities of FM operators, reconstructing the depth map can be challenging at times. However, the depth map's initial miscalculations can be balanced by adding some extra information about the shape of the object. This data is referred to as a guidance map or guidance image. Figure 4.5(a), (b), and (c) show the mean guidance image for cosine, sine, and cone, respectively.



(a) Initial depth map of cone image



(b) Initial depth map of cosine image



(c) Initial depth map of sine image

Figure: 4.2 Initial depth map of different images sequences

According to a recently proposed study, mean image guidance maps aid in the improvement of the initial depth map and produce superior results when compared to other guidance maps [36] can be seen in the Figure 4.3. The values obtained using the mean guidance map is shown in the first index in Figure 4.3. The mean image intensity guidance map along the image sequence's z axis can be calculate by,

$$G_{avgi}(u,v) = \frac{1}{N}\sum_{i=1}^{N} I_i(u,v) \quad (6)$$

where G_{avgi}, mean image intensity in the image sequence along z direction and N total number of pixels in the image, and the position of pixels are $I_i(u,v)$.

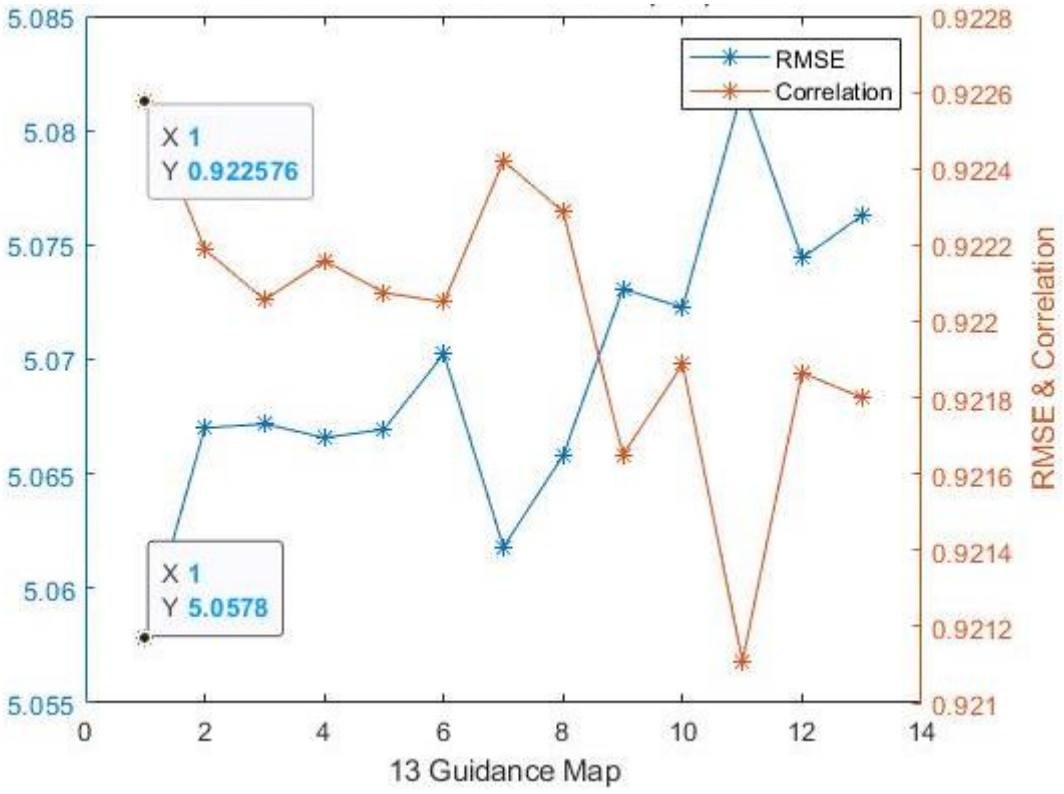


Figure: 4.3 RMSE and Correlation using 13 different guidance maps adopted from [36]

4.3 Guided Image Filters

The guided image filters update an image, using the information of another image, called a guidance map or guidance image. In this study, mean image is used as a guidance map, to impact the filtering. The guidance map is also an image itself, a different version of the scene. Guided image filtering works same as the other filtering operations works, but when calculating the value of the output pixel, the guided image considers the statistics of the

relevant local neighborhood area. Guided filters are mainly categorized into two classes based on number of inputs are implicated in the process of filtering. First category is self-guided in which one image is involved. In the second category, referenced-based and mutual guided where two images are involved in filtering process. This study proposed a framework for the enhancement of depth map using a weighted combination, where used different reference-based and mutual guided filters that are shown in Table 2.1. The general processing flow of guided filters is shown in Figure 4.4. Guided filters operate on two input images, one is the target image also known as input image and second is the guidance image. Guidance image is used as a reference to filter input image. The guidance map image we used in this study is an mean input image. The mean guidance image delivers better outcomes than alternative guidance maps [36]. The target image is an initial depth map that we aim to improve.

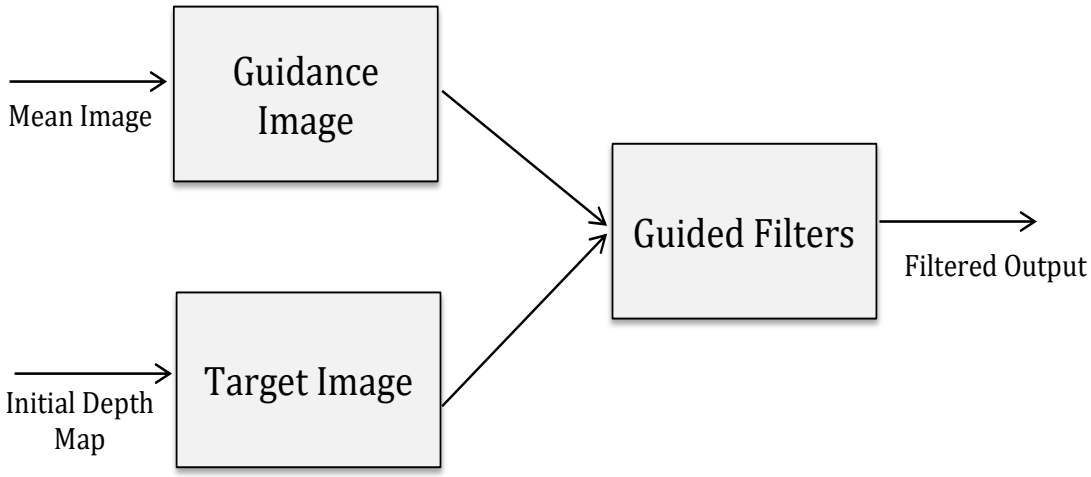
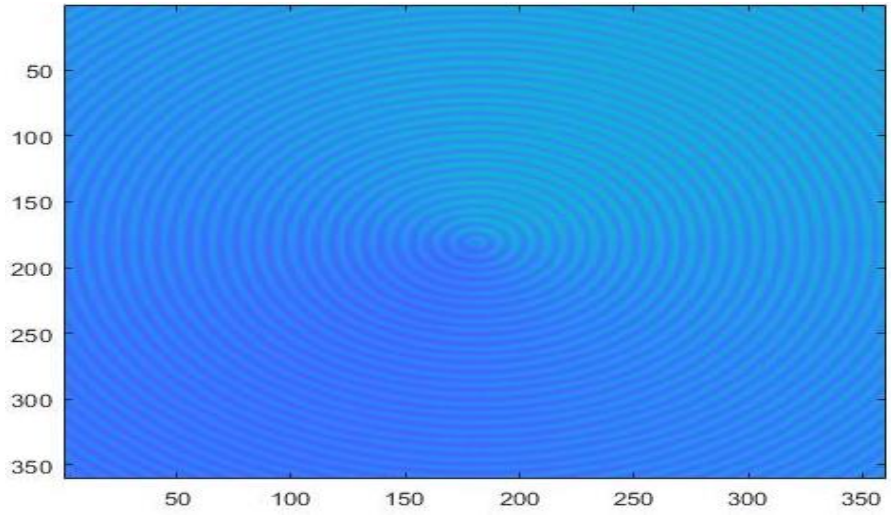


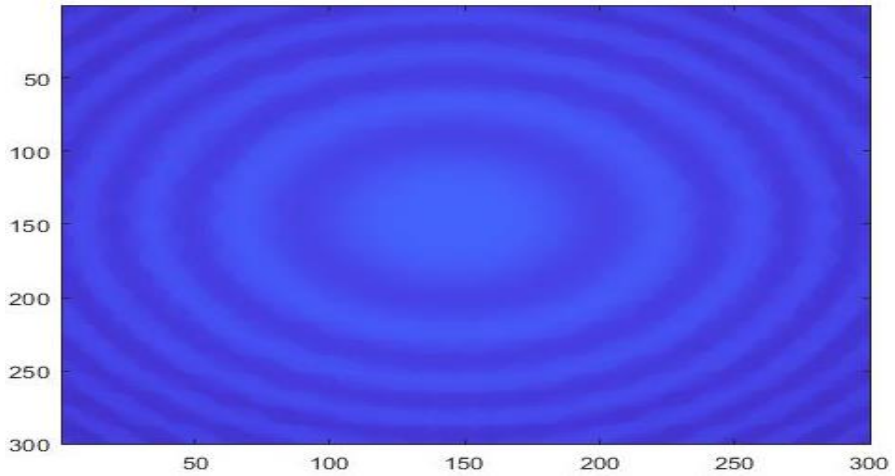
Figure: 4.4 Guided filters flowchart

4.4 Sequential Forward Search

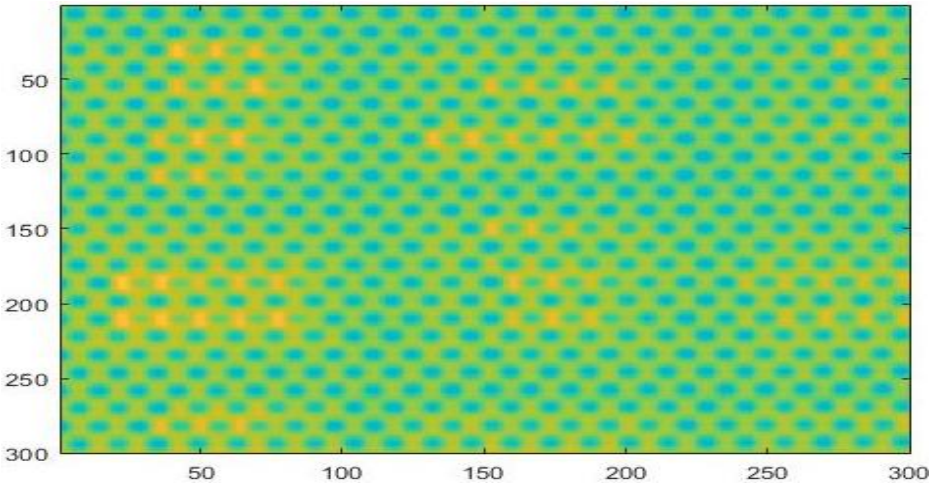
Sequential forward search algorithms are primarily a function of wrapper search methods [26]. Where it puts and removes sequential features from the problem area. In some cases, it explores each element independently and chooses M features from N features considering individual scores. This method is called naive feature selection. SFS approach is used in proposed framework to find best combination of filters. On the first iteration SFS techniques examine the results of 19 filters individually and on the second iteration it increases the set of filters and add the best performing filter into guided filter set and try all the remaining filter combination. Where found that a set of 2 filters is giving better results for



(a) Mean guidance image of cone image



(b) Mean guidance image of cosine image



(c) Mean guidance image of sine image

Figure: 4.5 Mean guidance image of different images sequences

depth map enhancement. When the performance stopped growing, the algorithm stops operating. In below Figure 4.6, general SFS flow chart is shown.

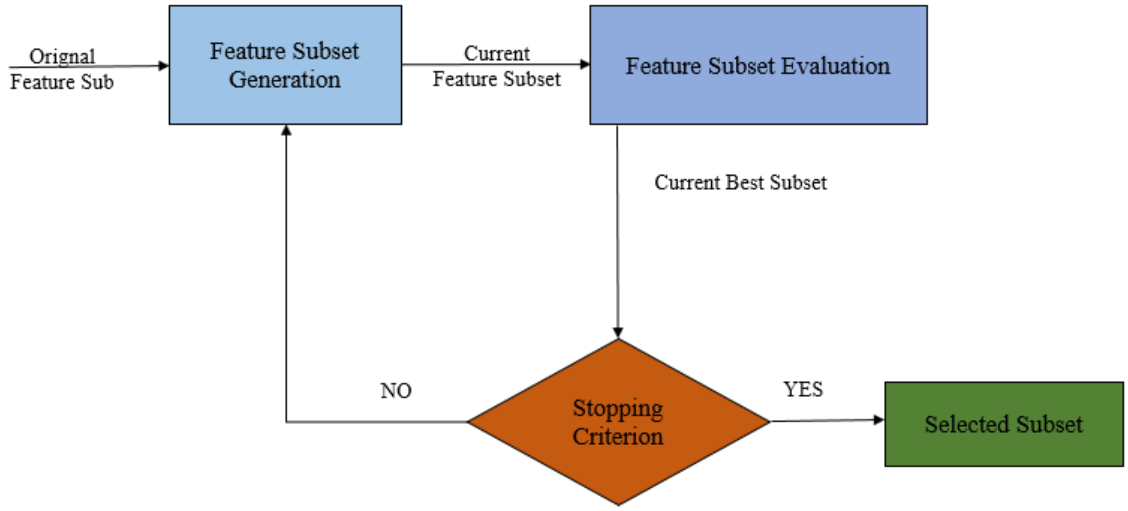


Figure: 4.6 General SFS flowchart

Table 4.1: Top 10 Filter for Cosine Image.

Filters	Set of 1F	Set of 2F	Set of 3F	Set of 4F	Set of 5F
CWMF	**7.778**	#	#	#	#
MRFF	7.8634	**7.2164**	#	#	#
FWLS	7.8642	7.2701	**7.2759**	#	#
GAD	7.8651	7.3047	7.2891	7.2932	7.3088
NAF	7.8656	7.291	7.2832	7.2957	7.3093
EGIF	7.8657	7.2511	7.2991	**7.2881**	#
JBU	7.8658	7.2639	7.2865	7.2946	7.3106
WGIF	7.8662	7.2524	7.2829	7.3029	**7.3054**
GIF	7.8662	7.2681	7.3097	7.311	7.3074
GGIF	7.8662	7.2788	7.3193	7.2962	7.3149

As indicated in the Table 4.1 below, we used the SFS technique to make combinations of 10 filters into a 1-filter subset to five-filter set. Initially, single filter applied on the initial depth map individually. When results of individual filters compared to the other filters at first iteration on cosine image, the EGIF filter improves the depth map the better than the others. As a result, the filter EGIF will make the combination with all the remaining filters in the next iteration. The hashtag "#" in the table indicates that the filter is active in the current iteration. In the second iteration, a set of two filters, the EGIF filters, combine with all other filters, and in the third iteration, the EGIF and MGDD filters combine with all other filters.

Table 4.2: Top 10 Filter for Cosine Image.

Filters	Set of 1F	Set of 2F	Set of 3F	Set of 4F	Set of 5F
EGIF	**4.1001**	#	#	#	#
MSJF	4.1287	4.1176	4.1034	4.1335	**4.1058**
FWMF	4.2072	4.1201	4.1086	**4.1047**	#
WGIF	4.2237	4.1002	4.1083	4.1061	4.1208
GGIF	4.2238	4.1059	4.1231	4.1069	4.1159
MGDD	4.2567	**4.0989**	#	#	#
GAD	4.2818	4.1012	4.1007	4.134	4.1234
FWL1	4.3085	4.1	**4.0998**	#	#
CWMF	4.3778	4.1202	4.1342	4.1421	4.1672
LSF	4.4165	4.1751	4.121	4.1178	4.126

For sinusoidal images, Table 4.2 shows SFS results for one to five sets of filter combinations where MRFF filters perform well in the first iteration. In the second iteration, MRFF and EGIF filters were used. For third iterations, the least RMSE filter combination is

MRFF, EGIF, and FWMF. In the fourth and fifth iterations of SFS with WGIF and LSF filters MRFF, EGIF, and FWMF filters provide the optimal combinational set of filters.

Table 4.3 illustrates SFS results for cone images with one to five sets of filter combinations where CWMF filters work effectively in the first iteration. In the second iteration, MRFF and CWMF filters have been used. The MRFF, CWMF, and FWLS filter combination with the lowest RMSE value in the third iteration. In the fourth and fifth iterations of SFS with EGIF and WGIF filters, MRFF, FWLS, and CWMF filters provide the appropriate combinational set of filters.

Table 2.3: Top 10 Filter for Sinusoidal Image.

Filters	Set of 1F	Set of 2F	Set of 3F	Set of 4F	Set of 5F
MRFF	**2.8909**	#	#	#	#
EGIF	2.8913	**2.861**	#	#	#
GAD	2.9038	2.8746	2.8665	2.8644	2.8687
MSJF	2.9062	2.8715	2.8655	2.8642	2.8605
FWMF	2.9068	2.8649	**2.8585**	#	#
LSF	2.9172	2.8814	2.8844	2.8696	2.8602
WGIF	2.9216	2.8787	2.8672	**2.863**	#
GGIF	2.9216	2.8789	2.8624	2.8727	2.8684
CWMF	2.9256	2.8638	2.8692	2.8666	**2.8588**
MGDD	2.9269	2.8823	2.8715	2.8681	2.8657

4.5 Particle Swarm Optimization

Particle swarm optimization (PSO) is a bio-inspired method that performs a basic search of the solution space for the optimal solution. It is distinguished from other

optimization techniques by the fact that it needs just the objective function and is unaffected by the objective's gradient or differential form. Additionally, it features a restricted set of hyperparameters.

Kennedy *et al.* [37] proposed particle swarm optimization approach in 1995. The systematic study of the biological believes that a group of fish or a flock of birds moving in a group "may profit from the experience of all other members". To put it another way, while a flock of birds is flying around looking for grain at random, all the birds in the flock can share their discoveries and help the entire flock have the best food possible.

Although we can copy the motion of the bird's flock. We can consider that each bird from the flock is helping us to find the best optimal result in a high-dimensional solution space. In this kind of approach, we can never fix a specific global optimal solution, called a heuristic approach. Although from the different experimental results it is discovered that the PSO approach solution is very close to the global optimal solution.

PSO can find better maxima or minima of the function on a multi-dimensional vector space. If we have a function f(X) that creates a real value from a vector parameter X(x,y) for this we can use PSO (where x, y are coordinate in a plane), and X can have almost any value in the space. For example, f(X) is the altitude, and we can get one value for any point on the plane. PSO approach will return you the X value that gives you the minima of function f(X).

Figure: 4.7 Scene of flock of birds

Let’s elaborate it with an example function:

$$f(x, y) = (y - 1.93)^2 + (x - \Pi)^2 + \sin(4x + 2.13) + \sin(6y - 2.31)$$

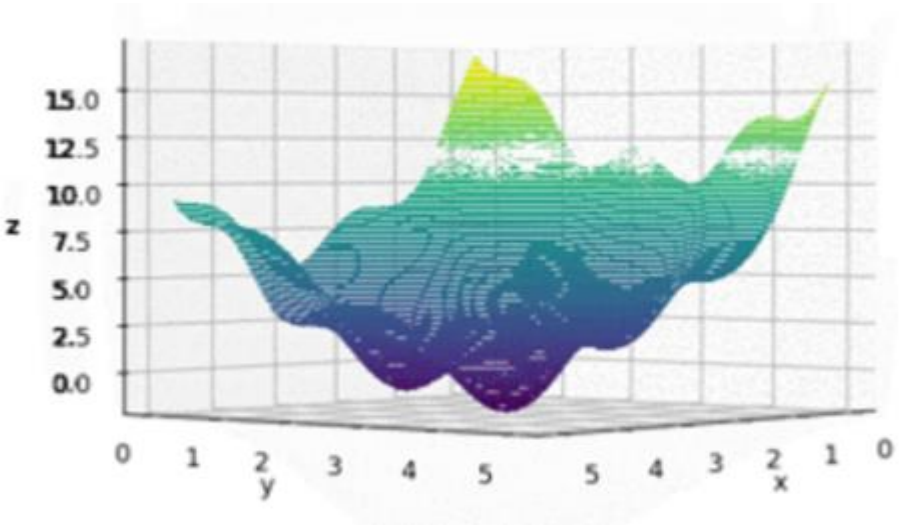


Figure: 4.8 Plot of above function f (x, y)

This function resembles a curved egg carton, as shown in the graph above. Because it is not a convex function, finding its minimum is difficult due to a local minima obtained is not all the time the global minima.

So, how to locate the function's minimal point? Certainly, we can conduct a thorough search: we can discover the least point by checking the value of f(x, y) for each location in the plane. If we believe it is too costly to search each point in the plane, may just locate few random examples points in the plane and determine which one gives the smallest value on f(x, y). However, we can see from the plot of f(x, y) Figure 4.8, that if we locate a point with a reduced f(x, y) value. It is easier to find a point with an even smaller value nearby.

A particle swarm optimization works in this way; we start with a bunch of random points in the solution space, called particles and let them look for the minima point in random paths, just like a bird’s flock looking for grain. At each step, each particle should seek around the lowest point it has ever discovered, as well as the smallest point discovered by the entire swarm of particles. After a certain number of cycles, we consider the function's minima point to be the lowest spot ever analyzed by this swarm of particles.

Assume there are P particles, and the location of particle k at iteration i is denoted by $X^k(i)$, which in the example above is a coordinate $X^k(i) = (x^k(i), y^k(i))$. Every particle has a velocity, which is represented by,

$$V^k(i) = v_x^k(i), v_y^k(i) \tag{7}$$

In addition to its position, the following is how the position of each particle would be changed in the next iteration.

$$X^k(i+1) = X^k(i) + V^k(i+1) \tag{8}$$

Or, homogeneously

$$X^k(i+1) = X^k(i) + V_x^k(i+1) \tag{9}$$

$$Y^k(i+1) = Y^k(i) + V_y^k(i+1) \tag{10}$$

The velocities are likewise adjusted by the rule at the same time:

$$V_k(i+1) = (w.V^k(i) + c_1 r_1(pbest - X^k(i)) + c_2 r_2(gbest - X^k(i) \tag{11}$$

where r_1 and r_2 are the random number ranging from 0 to 1. w, constants, c_1 and c_2 are PSO algorithm parameters. $pbest_k$ is the position that gives the best f(X) value ever explored by particle k and gbest is the position explored by all the particles in the swarm.

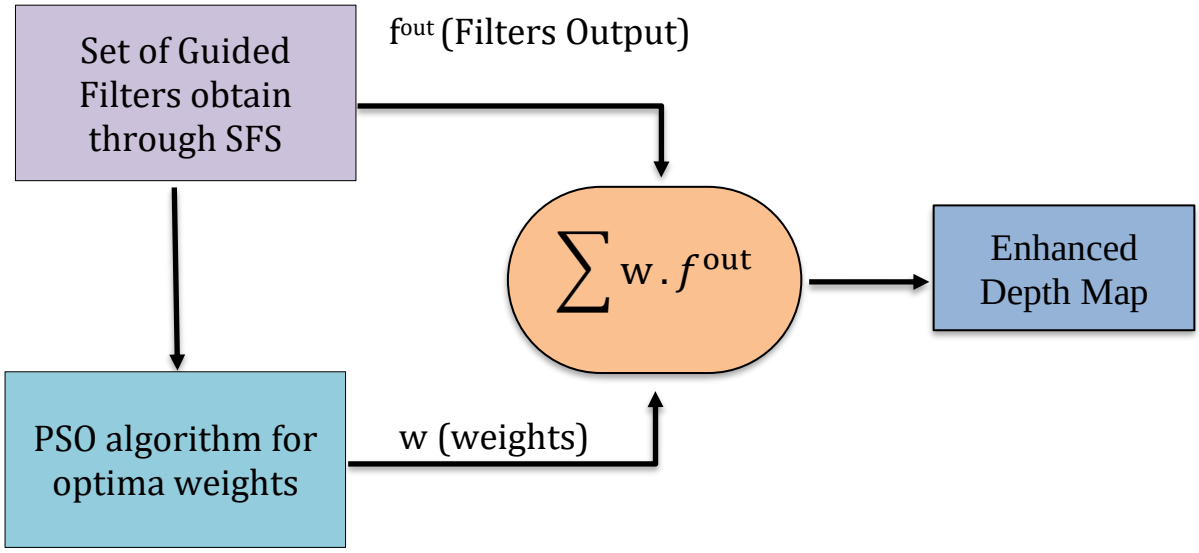


Figure: 4.9 Implementation of PSO algorithm and SFS approach

The parameter *w* is inertia weight. It has a value between 0 and 1 and specifies how much of the particle's previous velocity must be preserved (i.e., angle and speed of the search). The parameters c_1 and c_2 represents the cognitive and social coefficients. They determine how much emphasis should be placed on improving the particle's search result versus identifying the swarm's search result. And these variables can be thought of as governing the exploration-exploitation trade-off. Figure 4.10 displays the PSO cost optimization curve after 100 iterations to obtain optimum filter set weights.

In each cycle, the positions $pbest_i$ and gbest are modified to represent the best position ever discovered. In our study, the PSO approach is used to find the best weights for combinations of filters. These weights play a great role to enhance the depth map. When a set of the two best filters was achieved for increasing the depth map by sequential forward search, then two weights were required for each filter to make a weighted combination of filters to increase the depth map. As shown in below equation,

$$I_{enhanced\_depth\_map} = \sum_n w_n \cdot f_n^{out} \tag{12}$$

where *n* is the number of the filter in the best performing filter set and *w* are the weights obtained from PSO, ranging from 0.0 to 0.99 and f^{out} are the outputs of the filters from the filter set that are obtained from SFS.

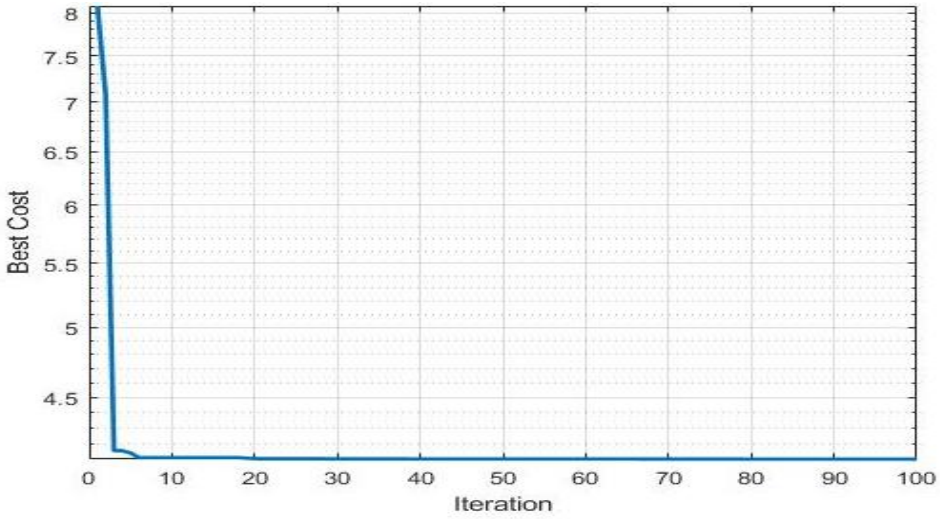


Figure: 4.10 PSO cost optimization plot

CHAPTER 5: EXPERIMENTAL RESULTS

In this section, we review the experimentally performance of the proposed framework. The proposed model was implemented in MATLAB ver.19a on a PC with 16GB RAM and a 3.30GHz processor. These filters need to set their parameters. In Section B, we experimentally tune all the filter parameters via extensive experimentation [36]. After that feeds these results to a sequential forward search to get the best combination of filters and the optimized weights for each filter are obtained from PSO. Finally, multiply the weights with the filter outputs that are chosen using SFS, and by adding these results we have a resultant image of the enhanced depth map.

5.1 Dataset

Synthetic objects have been used to assess the exhibit of guided image filters. The test dataset comprises three picture successions of manufactured articles or called synthetic objects. Synthetic objects are like artificial objects or shaped like cosine objects, con objects, and sinusoidal waves, whose depth map are produced using numerical functions. Gaussian estimation of the point spread capacity is processed utilizing the camera boundaries and the depth map of the item. More insight about picture models and engineered picture age can be found in [28]. The picture arrangement of the cone-shaped article has 97 pictures (dimension: 360 × 360). Sinusoidal images and cosine images arrangements comprise 60 pictures and each picture is of dimension 300 × 300 pixels. Sample images from synthetic image sequences are displayed in Figure 5.1.

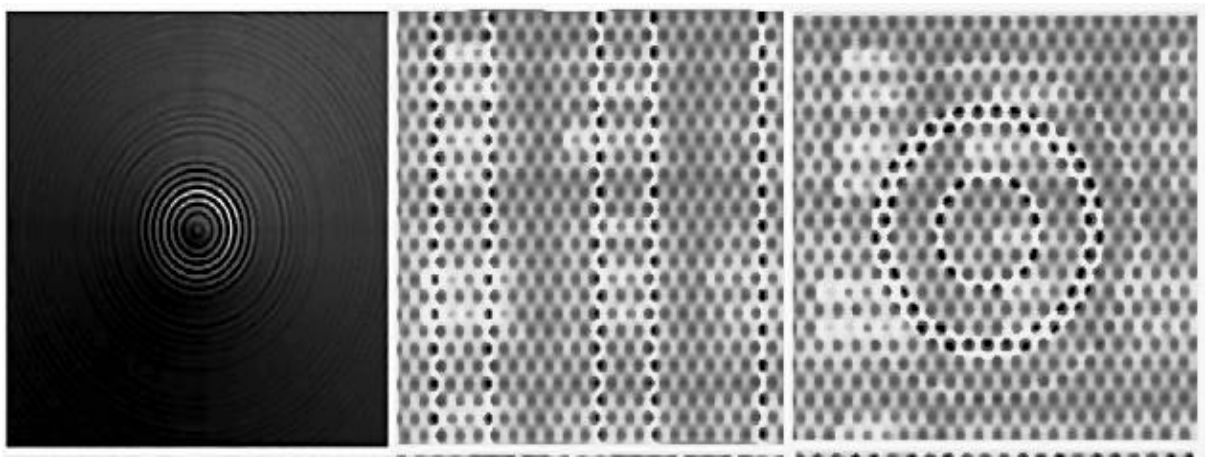


Figure: 5.1 Synthetic image sequences samples

5.2 Evaluation Measures

In order to evaluate the various guided filter implementations, it is necessary to use increased depth maps as a benchmark. Ground truth for synthetic image sequences is the actual depth map of these objects. Quantitative evaluation of the increased depth map's results is possible by comparing them to the ground truth. The root mean square blunder (RMSE) and the correlation are two quantitative measurements that are used (CORR). These two quantitative measures are given by,

$$RMSE = \sqrt{\frac{1}{XY}\sum_{u=1}^{X}\sum_{v=1}^{Y}\left(D_f(u,v) - D_g(u,v)\right)^2} \tag{13}$$

where $D_f(u,v)$, denote the filtered depth map and $D_g(u,v)$, ground truth depth map. While X and Y represent the total number of pixels in the map. And correlation coefficient given by,

$$\text{Correlation} = \frac{\sum_{u=1}^{X}\sum_{v=1}^{Y}\left(D_f(u,v)-D_m\right)\left(D(u,v)-D_g\right)}{\sqrt{\sum_{u=1}^{X}\sum_{v=1}^{Y}\left(D_f(u,v)-D_m\right)^2\sum_{u=1}^{X}\sum_{v=1}^{Y}\left(D(u,v)-D_g\right)^2}} \tag{14}$$

5.3 Optimal Parameter Tunning

In the proposed framework, weighted combinations of guided filters for the improvement of depth map in SFF. These guided filters are applied using open-source code made available by their authors. To achieve the highest possible average performance. The filter parameters were tuned experimentally, and their values are listed in Table 5.1. There are several parameters that are frequently used in these guided filters, including σ, indicates the Gaussian kernel's standard deviation, λ, indicates the regulation parameter, n, indicates the total number of iterations, and *w*, indicates the window's *w* x *w* size. The reader is referred to the pertinent literature [36] for a detailed explanation of these parameters.

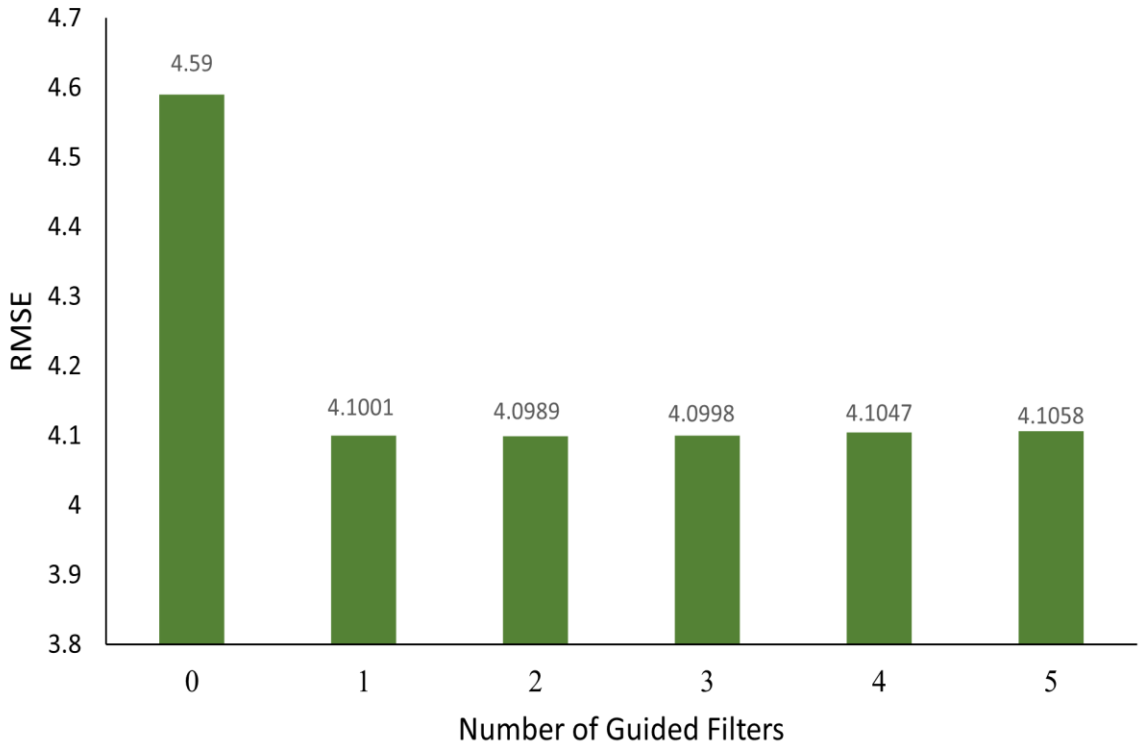
Table 3.1: Guided Filters Parameters Tuning [36].

No.	Filters	Filter Parameters	Sinusoidal Image	Cosine Image	Cone Image
1	NAF [26]	(σs, σr, σd, up, e, τ, w)	(0.2,10,2,1,0.5,20,7)	(2,5,5,1,0.1,5,5)	(0.2,10,2,1,0.5,20,7)
2	LSF [35]	(σr, λ)	(15,1)	(15,1)	(15,1)
3	JBF [21]	(σs, σr, w)	(1.2,0.25,5)	(1.2,0.25,5)	(1.2,0.25,5)
4	WMF [27]	(σr, σd, σs, w)	(10,3,3,21)	(16,6,2,17)	(16,6,2,17)
5	MRFF [34]	(σd, σs)	(0.9,0.1)	(0.1,0.3)	(1,1)
6	CWMF [29]	(λ)	(0.01)	(1102)	(1102)
7	GAD [28]	(σs, c)	(15,0.1)	(15,0.1)	(10,0.1)
8	MSIF [18]	(λD, λG, D, G, w, n)	(5x1010,10x1010,50,50,9,20)	(5x1010,10x1010,100,100,9,20)	(5x1010,10x1010,100,20,5,20)
9	FWL1 [24]	(e, n,σ, λ)	(5,5,0.02,50)	(5,5,0.01,0.5)	(5,5,0.01,0.9)
10	WGIF [31]	(λ, w)	(0.1,9)	(0.1,9)	(0.1,5)
11	FWLS [24]	(σ, e, n, λ)	(0.1,1.2,5,0.001)	(0.3,1.2,5,0.1)	(0.3,1.2,5,0.5)
12	SDF [20]	(λ, μ, nei, step, ν)	(2,20, 4,10,20)	(0.1,20,4,10,20)	(1,20, 8,10,10)

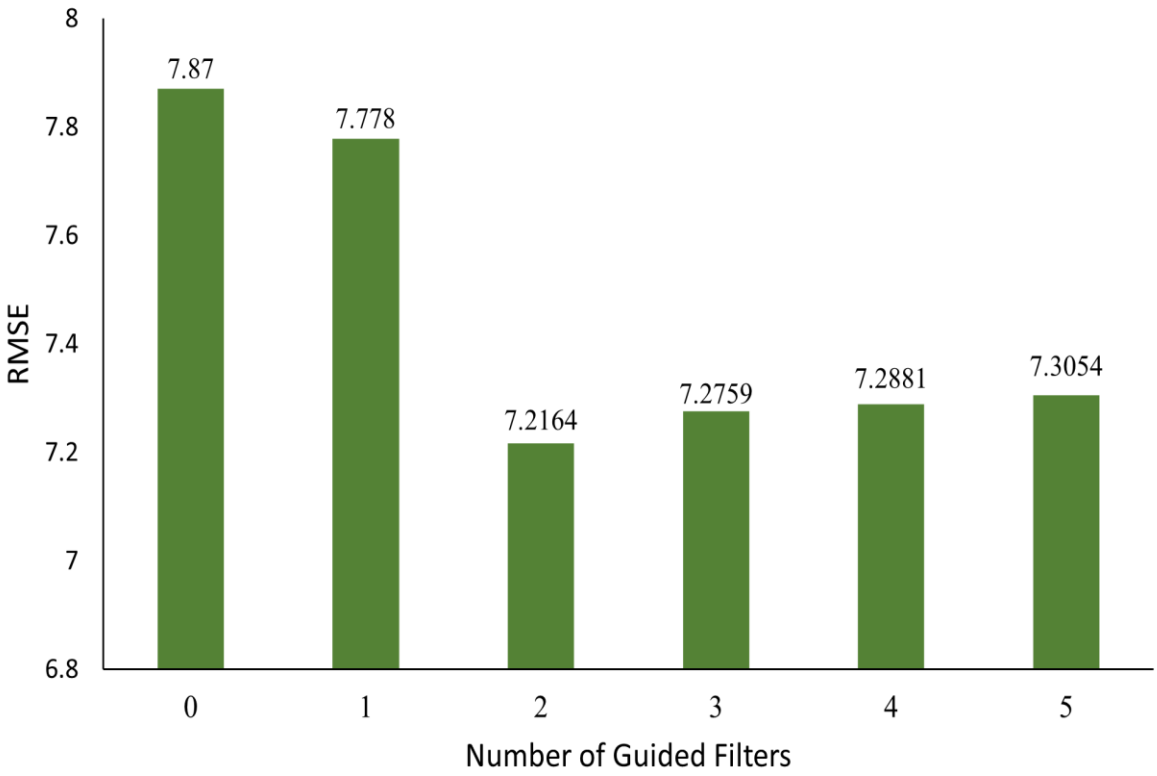
13	FWMF [30]	(σ, n, nT , nG, wt, w)	(100,1,256,256, exp,17)	(5,1,256,256,exp, 5)	(40,1,256,256, exp,21)
14	EGIF [133]	(λ, w)	(200,9)	(100,9)	(230,5)
15	GGIF [32]	(λ, w)	(0.01,9)	(0.01,9)	(0.01,5)
16	MGDD [19]	(n, λD, λG)	(10,0.1,90)	(10,0.01, 100)	(0.0005, 5,2)
17	JBU [25]	(σs, σr, up, w)	(0.5,0.1,1,5)	(0.5,0.1, 1,5)	(3,3,1,5)
18	MSDG [19]	(λD, n)	(0.1,10)	(0.001,10)	(0.001,10)
19	GIF [17]	(λ, w)	(0.0001,11)	(0.0001,11)	(0.0001,11)

5.4 Results and Discussion

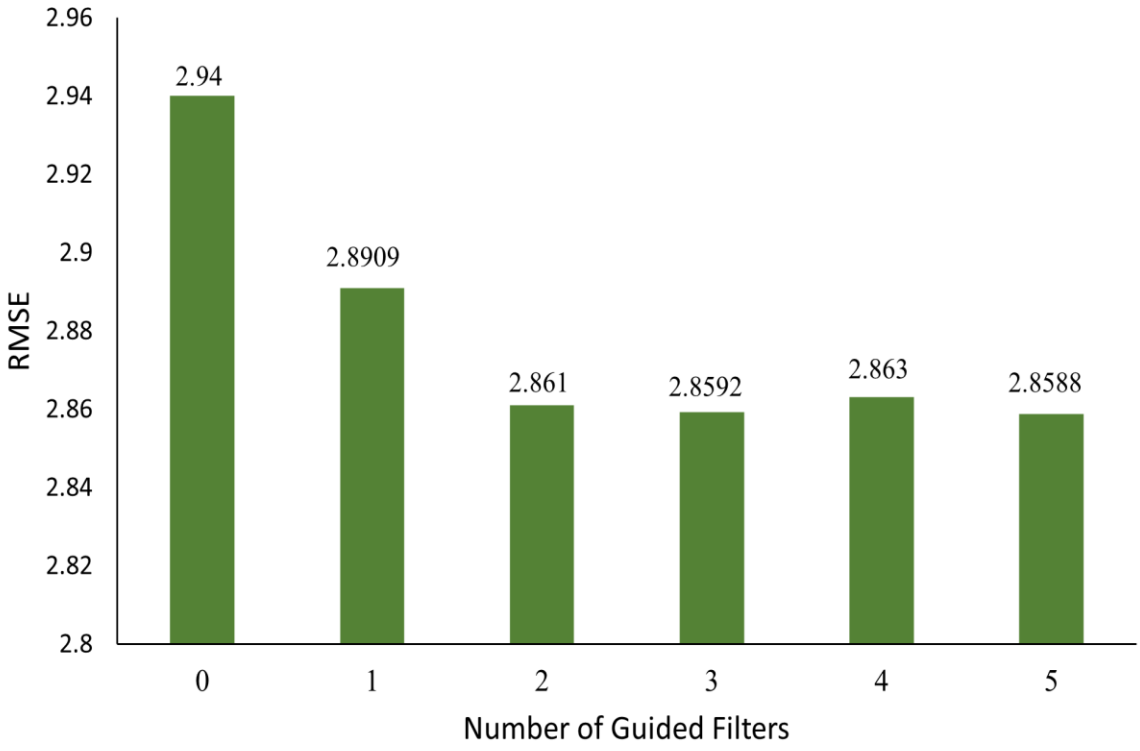
Initially, we scrutinize the quality of weighted combinations of guided filters, root mean square error quantitative measures have been shown in Figure 5.3, and Figure 5.4 for synthetic images. Depth maps of cosine image are shown in Figure 5.3(a). Below figure shows the initial depth map, and enhanced depth map using a weighted combination of guided filters from a single set of filters to a set of 5 filters. The depth map achieved by applying the weight combination of MSJF [18] and EGIF [33] filters (combination of 2 filters) enhanced than the initial depth maps.



(a) RMSE values for cosine sequence



(b) RMSE values for cone sequence



(c) RMSE values for sinusoidal sequence

Figure: 5.3 Root mean square error of enhanced depth map with respect to ground truth using multiple data sets and different combination of guided filters

The enhanced depth map for the cone images sequences are displayed in Figure 5.3(b) and it is observed that the weight combination of guided filters (MRFF [34] and CWMF [29] filters) have created an enhanced depth map with respect to the initial depth map. Sinusoidal object depth map is shown in Figure 5.3(c) and for the sinusoidal object set of 5 weighted combinations guided filters (LSF [35], FWMF [30], EGIF [33], MRFF [18], and CWMF [29] filters) give an enhanced depth map as compared to the initial depth map.

For a better performance review of weighted combinations of guided filters, all the results of the depth map shown in Figure 3 and the average results shown in Figure 5.4. It has been observed that a set of 2 filters gives better results than other filter combinations. As shown in Figure 5.4, the average RMSE for sinusoidal, cosine, and con image is lower at a combination of 2 filters than other filter combination sets. Where for cosine image, a weighted combination of MSJF [12] and EGIF [17] filter perform better, and for the sinusoidal image, it is observed that weighted combination of MRFF [13], and EGIF [17] filter perform well and for cone image MRFF [13], CWMF [9] filters weighted combination gives better performance.

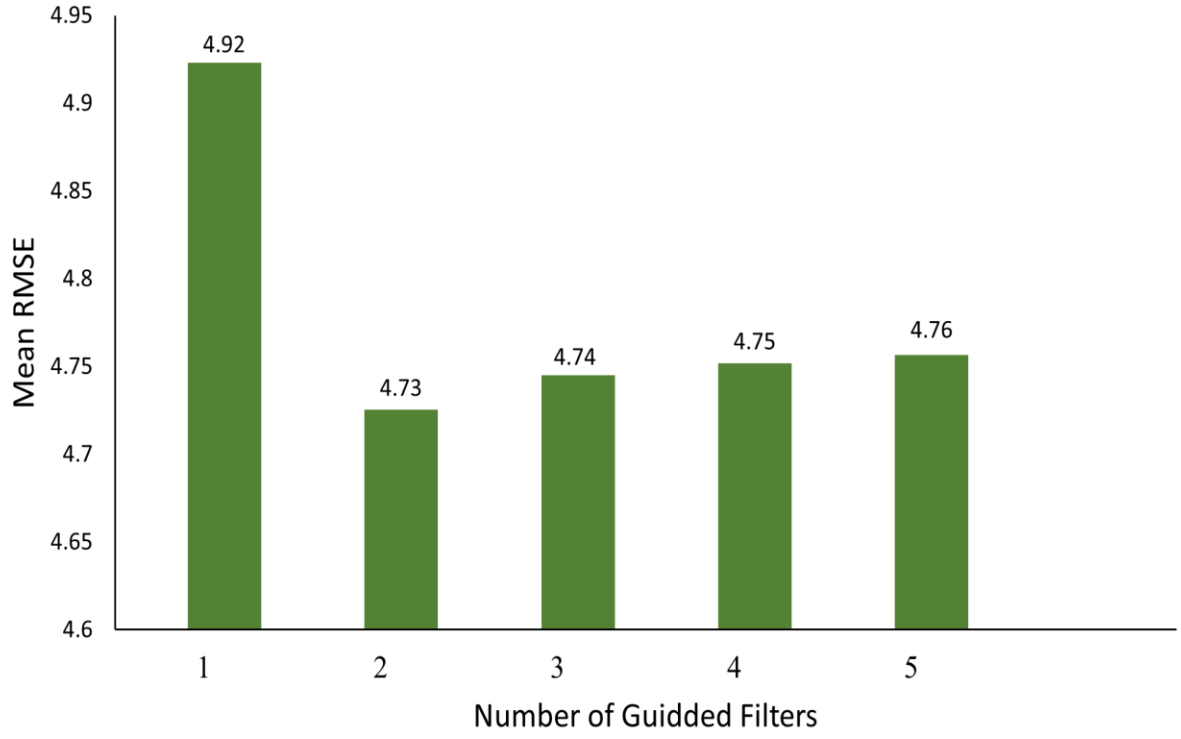


Figure: 5.4 Mean RMSE computed over three image sequences sine, cosine, and cone with respect to each filter set (0-5)

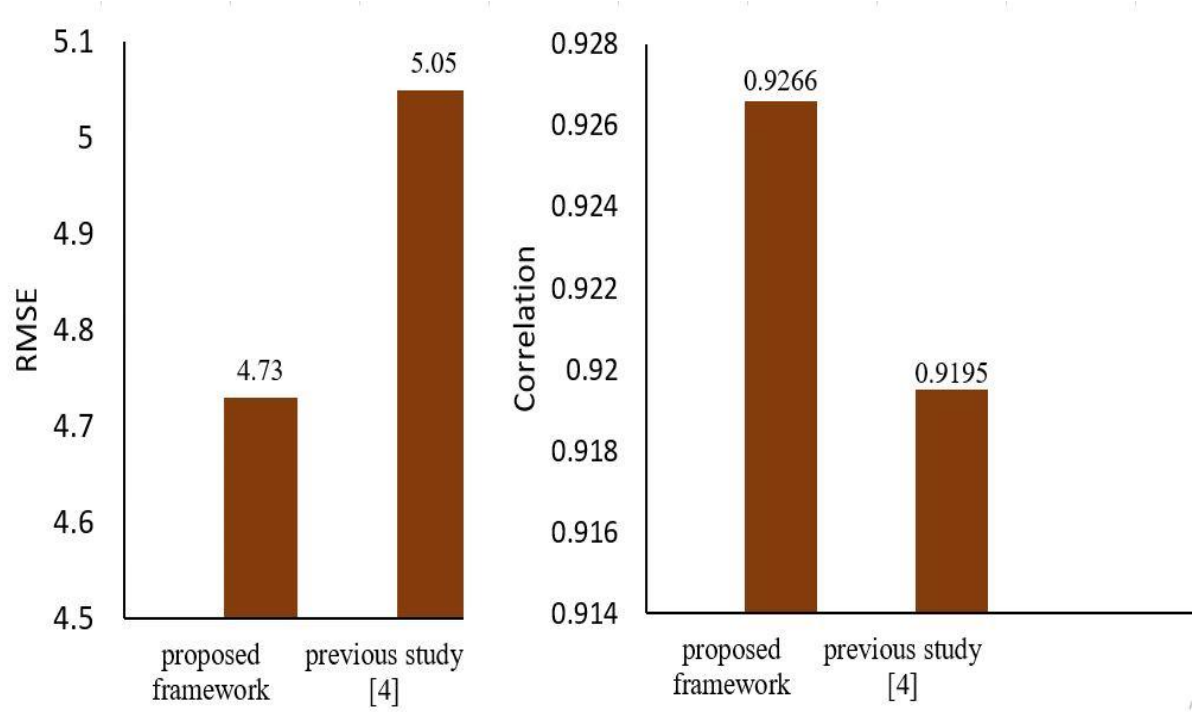


Figure: 5.5 Performance comparison, in terms of correlation and mean RMSE, between proposed approach and recent work by Usman *et al.* [36]

It has been observed that our proposed framework has improved results as compared to the initial depth map. The comparative results of the proposed framework and previous proposed study [36] are shown in Figure 5.5, where RMSE value is 4.7 and 93% correlation with respect to the actual depth map. Our proposed approach has been found to produce better results. only using a set of two filters with respect to less RMSE and more correlation to the actual depth map, while in a recently proposed study [36], 19 guided filters were included to enhance the depth map.

CHAPTER 6: CONCLUSION & FUTURE WORK

6.1 Conclusion

Shape-From-Focus uses focus in place of a cue to recreate the scene's 3D shape. FM operator's performance is influenced by a variety of elements including as picture illumination, consistency of a surface, distinguish objects colors, quantity of noise, and imaging equipment attributes. As a result, there is a chance that the rebuilt depth map contains some incorrect depth estimates. This research discovers that, the weighted combination subset of guided filters is useful and efficient in enhancing the depth image. The chosen guided filters have been picked after a thorough assessment of the most recent literature. Guided filtering has been found to be helpful in enhancing the initial depth map image. We assess different image filters in the study, and they all offer ways to enhance the initial depth map. The proposed framework utilizes two guided filters subset that leads to fewer computations and enhanced depth maps as compared to earlier works.

In conclusion, the findings of this study show how different weighted combination of guided image filters improve depth image. We feel that our research on depth map enhancement is important and that it will lead to future progress in the field of SFF using guided filtering.

6.2 Contribution

- A novel framework for the enhancement of depth map using guided image filters in shape-from-focus.
- In the depth enhancement framework, the optimized set of guided image filters has been proposed.
- Recent breakthroughs in depth map enhancement are reviewed and compared.

6.3 Future Work

In the future, a robust guided image filter could be used to capture image details to enhance the depth map. A unique guidance map image has been suggested which helps in guided image filtering or any other method of weight optimization that can be used to improve the depth map.

REFERENCES

[1] P. Thanusutiyabhorn, P. Kanongchaiyos, W.S. Mohammed, Image-based 3d laserscanner, in: The 8th Electr. Engg./Electronics, Computer, Telecomm. and Information Technology, IEEE, 2011, pp. 975–978.

[2] S.K. Nayar, Y. Nakagawa, Shape from focus, IEEE Trans. Pattern Anal. Mach. Intell. 16 (8) (1994) 824–831.

[3] M.T. Mahmood, T.-S. Choi, Nonlinear approach for enhancement of image focus volume in shape from focus, IEEE Trans. Image Process. 21 (5) (2012) 2866–2873.

[4] Alicona, Optical measurement solutions in use, accessed on Dec. 10, 2019. https://www.alicona.com/applications/.

[5] S. Suwajanakorn, C. Hernandez, S.M. Seitz, Depth from focus with your mobile phone, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3497–3506.

[6] J. Surh, H.-G. Jeon, Y. Park, S. Im, H. Ha, I. So Kweon, Noise robust depth from focus using a ring difference filter, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6328–6337.

[7] S. Pertuz, D. Puig, M.A. Garcia, Analysis of focus measure operators for shape-fromfocus, Pattern Recognit. 46 (5) (2013) 1415–1432.

[8] A.S. Malik, T.-S. Choi, Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3d shape recovery, Pattern Recognit. 40 (1) (2007) 154–170.

[9] A. Thelen, S. Frey, S. Hirsch, P. Hering, Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood– size, and height value interpolation, IEEE Trans. Image Process. 18 (1) (2009) 151–157.

[10] U. Ali, V. Pruks, M.T. Mahmood, Image focus volume regularization for shape from focus through 3d weighted least squares, Inf. Sci. 489 (2019) 155–166.

[11] U. Ali, M.T. Mahmood, 3d shape recovery by aggregating 3d wavelet transformbased image focus volumes through 3d weighted least squares, J. Math Imag. Vis. 62 (1) (2020) 54–72.

[12] C. Ribal, N. Lermé, S. Le Hégarat-Mascle, Efficient graph cut optimization for shape from focus, J. Vis. Commun Image Represent. 55 (2018) 529–539.

[13] Z. Ma, D. Kim, Y.-G. Shin, Shape-from-focus reconstruction using nonlocal matting 42 laplacian prior followed by MRF-based refinement, Pattern Recognit. (2020) 107302.

[14] R. Minhas, A.A. Mohammed, Q.M.J. Wu, M.A. Sid-Ahmed, 3d shape from focus and depth map computation using steerable filters, in: International Conference Image Analysis and Recognition, Springer, 2009, pp. 573–583.

[15] V. Gaganov, A. Ignatenko, Robust shape from focus via markov random fields, in: Proceedings of Graphicon Conference, 2009, pp. 74–80.

[16] M. Moeller, M. Benning, C. Schönlieb, D. Cremers, Variational depth from focus reconstruction, IEEE Trans. Image Process. 24 (12) (2015) 5369–5378.

[17] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1397–1409.

[18] X. Shen, C. Zhou, L. Xu, J. Jia, Mutual-structure for joint filtering, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3406–3414.

[19] X. Guo, Y. Li, J. Ma, H. Ling, Mutually guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 42 (3) (2020) 694–707.

[20] B. Ham, M. Cho, J. Ponce, Robust guided image filtering using nonconvex potentials, IEEE Trans. Pattern Anal. Mach. Intell. 40 (1) (2018) 192–207.

[21] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, K. Toyama, Digital photography with flash and no-flash image pairs, ACM Trans. Graph. (TOG) 23 (3) (2004) 664–672.

[22] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: ICCV, 1998, pp. 839–846.

[23] Q. Zhang, X. Shen, L. Xu, J. Jia, Rolling guidance filter, in: European Conference on Computer Vision, Springer, 2014, pp. 815–830.

[24] Y. Kim, D. Min, B. Ham, K. Sohn, Fast domain decomposition for global image smoothing, IEEE Trans. Image Process. 26 (8) (2017) 4079–4091.

[25] J. Kopf, M.F. Cohen, D. Lischinski, M. Uyttendaele, Joint bilateral upsampling, ACM Trans. Graph. (ToG) 26 (3) (2007) 96–es.

[26] [26] D. Chan, H. Buisman, C. Theobalt, S. Thrun, A noise-aware filter for real-time depth upsampling, Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, 2008.

[27] D. Min, J. Lu, M.N. Do, Depth video enhancement based on weighted mode filtering, IEEE Trans. Image Process. 21 (3) (2012) 1176–1190.

[28] J. Liu, X. Gong, Guided depth enhancement via anisotropic diffusion, in: Advances in Multimedia Information 43 Processing, volume 8294, Springer, 2013, pp. 408–417.

[29] Z. Ma, K. He, Y. Wei, J. Sun, E. Wu, Constant time weighted median filtering for stereo matching and beyond, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 49–56.

[30] Q. Zhang, L. Xu, J. Jia, 100+ times faster weighted median filter (WMF), in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2830–2837.

[31] Z. Li, J. Zheng, Z. Zhu, W. Yao, S. Wu, Weighted guided image filtering, IEEE Trans. Image Process. 24 (1) (2015) 120–129.

[32] F. Kou, W. Chen, C. Wen, Z. Li, Gradient domain guided image filtering, IEEE Trans. Image Process. 24 (11) (2015) 4528–4539.

[33] Z. Lu, B. Long, K. Li, F. Lu, Effective guided image filtering for contrast enhancement, IEEE Signal Process. Lett. 25 (10) (2018) 1585–1589.

[34] J. Diebel, S. Thrun, An application of markov random fields to range sensing, in: Advances in neural information processing systems, 2005, pp. 291–298.

[35] A. Harrison, P. Newman, Image and sparse laser fusion for dense scene reconstruction, in: Field and Service Robotics, Springer, 2010, pp. 219–228.

[36] U. Ali, I. H. Lee, and M. T. Mahmood, “Guided image filtering in shapefrom-focus: A comparative analysis,” Pattern Recognition, vol. 111, p. 107670, 2021.

[37] U. Ali and M. T. Mahmood, “Energy minimization for image focus volume in shape from focus,” Pattern Recognition, vol. 126, p. 108559,2022.

[38] U. Ali and M. T. Mahmood, “Robust focus volume regularization in shape from focus,” IEEE Transactions on Image Processing, vol. 30, pp. 7215–7227, 2021.

[39] P. Pudil, J. Novovicov ˇ a, and J. Kittler, “Floating search methods in ´ feature selection,” Pattern recognition letters, vol. 15, no. 11, pp. 1119– 1125, 1994.

[40] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in Proceedings of ICNN’95-international conference on neural networks, vol. 4, pp. 1942–1948, IEEE, 1995.