1-1-2011

Three-dimensional filters for multiview stereoscopic applications using layered depth images

Alexander S. Babalis
Ryerson University

THREE-DIMENSIONAL FILTERS FOR MULTIVIEW STEREOSCOPIC APPLICATIONS USING LAYERED DEPTH IMAGES

by

Alexander S. Babalis, B.A.Sc.
University of Toronto, 2009

A thesis
presented to Ryerson University
in partial fulfillment of the
requirements for the degree of
Master of Applied Science
in the Program of
Electrical and Computer Engineering

Toronto, Ontario, Canada, 2011

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature

I further authorize Ryerson 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.

Signature

Instructions on Borrowers

Ryerson University requires the signatures of all persons using or photocopying this thesis.

Please sign below, and give address and date.

Abstract

THREE-DIMENSIONAL FILTERS FOR MULTIVIEW STEREOSCOPIC APPLICATIONS USING LAYERED DEPTH IMAGES

Alexander S. Babalis
Master of Applied Science
Department of Electrical and Computer Engineering
Ryerson University
2011

This thesis proposes an extension of two-dimensional (2D) spatial filtering into three-dimensions for multiview stereoscopic applications using the layered depth image (LDI) representation. The proposed filtering scheme takes advantage of the depth information available when an image is represented with layers, and can return results that are comparable to or better than 2D filtering techniques for smoothing or sharpening stereoscopic images. In addition, the proposed filtering scheme is more efficient for multiview stereoscopic applications using LDIs than conventional 2D filtering since the filter needs to be applied only once for n views, whereas 2D filtering requires each view to be filtered separately (increasing computation time).

The proposed filtering method for smoothing stereoscopic images was also subjectively evaluated in a study involving 15 people. The results from this study indicated that the proposed filtering scheme received similar scores for both viewer comfort and naturalness when compared to the 2D bilateral filter.

Acknowledgements

I first and foremost want to thank my family for supporting me from day one. Mom, Dad, and Joanne, I realize now how valuable my education has been, and I will be forever indebted to you for being there for me throughout and for encouraging me to pursue graduate studies. You each have helped me get through the toughest of times and always encouraged me to reach for the top in everything I do. In addition, I also want to mention and thank my father for being the original person who made me interested in stereoscopic 3D, and for motivating me to conduct research in this emerging area. I also want to thank the rest of my extended family (too many people to name), but especially my cousin Maria for her help in this work. Lastly, I want to thank my girlfriend Laura who helped me get through each day, and for putting up with my dedication to research even when it required me to invest extra hours over the weekend.

I want to thank Dr. Venetsanopoulos, my supervisor, for having the confidence in me and taking me as a summer student in 2009. Since then, you have nurtured me, inspired me, and taught me a tremendous amount about how to conduct research. I want to sincerely thank you for giving me this opportunity to shine and helping me grow my knowledge base in signal and image processing. The continuous guidance and endless support, related to this thesis and other matters, will never be forgotten. Even during your busiest of times as Vice-President, Research and Innovation at Ryerson University, you still managed (several times a week) to find time for each of your students. Without you, this thesis wouldn't have been possible.

I also want to sincerely thank Dr. Androutsos, my co-supervisor, for providing me with countless feedback and helping to steer me in the right direction throughout the course of my degree. Your expertise in the area of stereoscopic imaging helped shape my thesis from the time of selecting a topic, leading up (and through) to completion. The recommendations that you made have been integral to my success. I am truly grateful for all of your time, advice and everlasting support.

Another person whom I would like to thank is Samuel Zhou. You were the first person who really taught me about stereoscopic 3D, and about the technology that is state of the art. You helped to shed light on the impact stereoscopic 3D is having in the industry and fuelled my passion for this area. The time you took out of your busy schedule to consult me during the early stages of my research facilitated my understanding of many important topics in the area that remained unsolved or that had yet to be explored.

Credit must also be given to the faculty in the Department of Electrical and Computer Engineering at Ryerson University who taught my graduate level courses. I also want to particularly thank the members of my examining committee for providing me with key feedback regarding my thesis.

Furthermore, I also want to extend my genuine gratitude and appreciation to the members (who I consider as friends) of the *Multimedia and Distributed Computing lab* and the *Digital Media Lab* at Ryerson that have come and gone: Farhang, April, Obaidul, Richard, Ray, Moe, Yasir, and Ricardo. You have all helped me in your own ways, but above all, thank you all for making my time here enjoyable. I want to thank you all for your kind support, guidance and your help when requested.

My gratitude must additionally be paid to the rest of my close friends. You guys were always there when I needed you and I could always count on you when I needed advice. I thank each of you for helping me stay positive throughout the entire length of my program and helping me escape to reality during my personal time.

Finally, I want to deeply thank Microsoft Research for sharing their work (*MSR 3D Video download*), which allowed me to have a medium to test my results. I sincerely thank you, and every other researcher in the scientific community, who makes their work publicly available, which in turn enables us all to collectively move forward.

To my parents.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

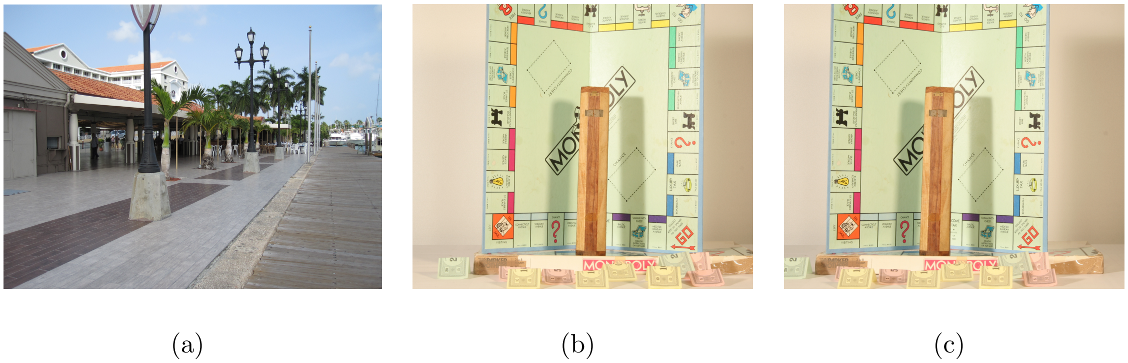
We are always seeing transitions.

We have seen media change before our eyes on many occasions. Technology like video cassettes changing to DVDs, and then to Blu-ray. Likewise, the way we reproduce images onto a display device has also evolved. We have seen television (TV) change from black and white, to colour, and from standard definition, to high-definition (HD). Now, we are seeing the transition of HDTV to stereoscopic (3D) TV (3DTV).

Stereoscopic imaging has also been evolving within itself. We are seeing more systems move from traditional forms of single view stereoscopic video (also known as conventional stereo) to multiple stereoscopic views or multiview stereo. Likewise, we are even seeing promising signs for the possible future transition to volumetric systems for 3D given recent successes in research [1].

1.1 Stereoscopic Imaging

Stereoscopic imaging and 3DTV is a more realistic form of image reproduction than 2D TV since it stimulates both eyes of the human visual system, similarly to the natural way one views the world. By seeing two different images (from slightly different angles), the viewer is able to perceive depth, known as *stereopsis*, which is different from monocular depth seen/portrayed in 2D images (see Fig. 1.1).



(a) (b) (c)

Figure 1.1: Comparison of monocular depth and binocular depth. (a) Monocular depth cues (ex. converging parallel lines, texture gradients) from a single image. (b,c) 2D images of the same scene taken from different angles. When viewed separately by each eye, the viewer perceives depth. Images (b,c) are different views of the Middlebury stereo dataset 'Monopoly' as mentioned in [2].

Since the renaissance of 3D cinema in 2009, catalized by advances made in digital cinema, stereoscopic 3D has become a very promising technology for the near and distant future. With the ever growing success of 3D movies at the box office, 3DTV in the home has yet to really take off for various reasons including lack of content, uncertainty regarding the technology (i.e. is it a fad?) and having to wear special glasses.

Ongoing research is actively taking place in the area of stereoscopic image/video processing to solve/reduce some of these problems associated with the display device. Namely, the transition to glasses-free displays is surfacing with autostereoscopic displays being produced with increased resolution. However, a fundamental problem that restricts the ability of current screens to appear more realistic to the consumer is the restriction to a single view of a 3D scene. To put it more clearly, the current state of 3DTV (or conventional stereo video) shows a viewer the same perspective of an image regardless of their viewing position, which is counter intuitive to what one would expect from a "life-like" technology. Specifically, if the viewer moves their head (also known as head motion parallax [3]), they would expect the scene to change as it would in real life, but this is not the case.

To grant the viewer with visual freedom, the problem of a single view can be solved through the extension of stereoscopic images into multiple stereoscopic views. One such solution to the challenge of multiple stereoscopic views is the concept of multiview video.

1.2 Multiview Imaging

In multiview imaging [4], the viewer is able to see the scene from more than one viewpoint (i.e. multiple views are available). This is achieved by capturing the original scene with multiple cameras from different angles and encoding each view into a single stream as in [H.264/MPEG-2 AVC]. When presented to the viewer, the stream can be used to render arbitrary locations (even where a camera did not exist) through interpolation techniques [5]. Ultimately, presenting the user with more views of the scene makes for a (potentially) more realistic 3D experience. Television manufacturers are looking toward multiview imaging for stereoscopic displays to give viewers freedom to move their heads and see the image change perspectives as it would happen in real life. Subsequently, multiview stereo has been dubbed the next, more immersive, form of stereoscopic imaging before other advanced technologies such as true holography are adopted [6].

Multiview video has a range of applications, in addition to multiview stereoscopic TV (which are applicable to both 2D, and stereoscopic imaging). Two examples of such applications are *Free Viewpoint Television* and *Virtual Museums*. Each of these applications are described below for the reader to have a better sense of how multiview video is applicable and beneficial to these applications.

- *Free Viewpoint Television:* Television, as we know it, is a one-way medium. The screen displays a 2D image, and the viewer watches. When viewing your favourite show, you do not have the ability to change the perspective from which you view the scene. To combat this limitation, a video format known as free view point television will allow the user to move around and chose from where they wish to view the scene [7]. This idea of a free-view can be accomplished using the concept of multiview video.

- *Art:* The preservation, admiration, and the study of 3D art objects, such as statues, pottery, and other antiques pose a challenge for museums. One way that this 3D art can be easily accessible for research purposes, or for admiration by those who are unable to visit the exhibit, is through the idea of virtual museums [8]. Virtual museums replicate 3D art on a monitor using 2D, or 3D technology. To make the virtual museum more realistic, multiview stereo technology can be adopted [9] such that the viewer can easily move around the scene and view the art from different angles. Having a system that is multiview capable, and one that produces stereo images can increase the realism of such art.

Multiview stereo can be achieved in an efficient manner using the concept of *layered depth video* which was derived from layered depth images [10]. A layered depth image (LDI) is an image representation for a three-dimensional scene using an array of two-dimensional image slices (also known as layers or an array of layered depth pixels as defined in [10]).

The benefit of using the LDI representation for multiview images [11] arises when one wishes to render a view of the multiview data from a different nearby viewpoint since the pixel information from the various layers can become visible to fill in regions that would have otherwise been occluded if *Depth-Image Based Rendering* (DIBR) was used [12]. Layered Depth Video also has the advantage of being a more compact representation than multiview video plus depth [13], [14] while also being invariant to the display type [15].

1.3 Image Processing for Multiview Stereo

Even though multiview stereo leads to a more realistic form of 3D imaging, it is still displayed on a 2D screen, and thus, many of the operations taking place from image acquisition until the display rely on 2D image processing. Moreover, many of the elementary image processing operations that have been successful in digital imaging to this day have predominantly been formulated for 2D images, and therefore do not consider the depth of the image. In order for stereoscopic and multiview imaging to be successful for the future in areas highly dependent on signal processing [6] (such as compression, noise reduction, and enhancement), an extension of the many 2D image processing operations into 3D is necessary for results to become more natural. One such image operation that has yet to be formally addressed in the context of stereoscopic images is filtering.

Image filtering is used throughout the field of image processing, whether for noise reduction, enhancement, or analysis of images. It is therefore very important for these types of operations to be carried over into 3D for stereoscopic images and future technologies such as multiview stereo.

1.4 Past Research in Filtering Stereoscopic Images

Numerous stereoscopic image processing techniques exist, but few of them have considered how to effectively perform spatial filtering on a stereoscopic image. Gonzalez and Woods [16] have described 2D spatial filtering, but filtering a stereoscopic image by applying a 2D spatial filter to the left and right images can lead to undesirable results. For example, using a smoothing filter on left and right images separately leads to blurring of the image components between the background and foreground. This is an undesirable effect that occurs with stereoscopic images that can lead to the viewer not having a clear distinction of the foreground and background.

Morgenthaler described three-dimensional digital image filtering in [17], but did not consider its effects in a stereoscopic image sense, and rather focused on medical images.

Papari *et al* [18] proposed a method to generate artistic stereo images by applying an edge preserving smoothing filter to left and right images. The method used the depth of objects to determine the degree of filtering. The work assumed that closer objects have more detail then those that were further away. In addition, the work did not perform any human subjective testing to support the method's ability to preserve the stereo effect.

Horng *et al* [19] applied a directional Gaussian filter to reduce holes in conventional (single view) stereoscopic images generated from DIBR while maintaining the perceived depth quality. The resulting images were evaluated by human subjects and the conclusions were that the proposed method could obtain warped stereoscopic images with good perceived naturalness and depth quality. Computation time, or how many iterations it took to arrive at the results were not mentioned.

Another area where filtering has been applied to stereoscopic imaging is in post-processing of images to aid in reducing viewer discomfort. Specifically, even though multiview stereo leads to a more realistic form of 3D imaging, it is still displayed on a 2D screen, which will lead to viewing discomfort.

1.4.1 Filtering for Reducing Viewer Discomfort

The problems associated with multiview images being shown on a screen are similar to those encountered by conventional (single view) stereo. One such problem is that the images displayed on the screen appear unnaturally sharp across all depths (which contradicts the basic nature of the human visual system) [20]. For example, when our eyes are focused on a certain object in space, our eyes accomodate (the lens inside our eye focuses on the point in space) while our eye moves inwards or outwards to locate the point (known as vergence). When this happens, after a certain limit (set by Panum's fusional area) all points beyond or in front of this region of focus appear blurred or as double images. An example of this type of scenario is shown in Fig. 1.4.1.

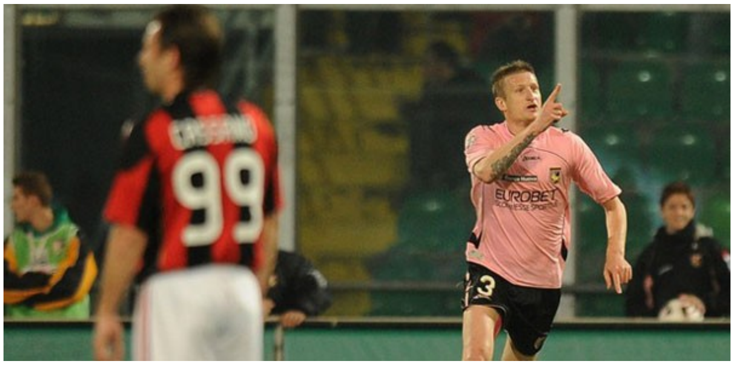


Figure 1.2: Example of a typical image seen with depth of focus. The soccer player slightly to the right from the center of the image appears sharp while the player infront and the spectators further behind appear blurred. The image was obtained from ESPN Internet Ventures 2011: <http://soccernet.espn.go.com/report/_/id/304728?cc=5901>.

On the other hand, when looking at a computer display, every pixel appears in focus. The vergence of our eyes remains fixed onto the display, while the accomodation in our eyes adjusts to locate the points of focus, leaving both the vergence and accomodation out of sync. This notion of the differing accomodation and vergence when viewing a stereoscopic image on a 2D screen can lead to visual discomfort [20].

Various post-processing techniques have been developed to mitigate the symptoms caused by viewing unnaturally sharp images.

In computer graphics, Blohm *et al* [21] developed a synthetic depth of field for computer images using a lens model to replicate the human eye. This method has shown to reduce viewing discomfort and aid in making computer generated images appear more realistic. Shinya [22] developed a ray distribution technique to defocus other layers for computer graphics.

In the case of natural images, Gi Mun Um *et al* [23] developed disparity-based asymmetrical filtering by blurring a single image to reduce discomfort, while still maintaining image quality. This method was applied in a conventional (single view) stereoscopic context. A previous study also revealed that blurring a single image did not impact the human visual perception [24].

Liu *et al* [25] developed a system comprised of three cameras to acquire images from a natural scene to perform color segmentation and stereo matching to compute disparity maps. After segmentation, the regions would be arranged into layers where the layered depth representation was used to render intermediate views (for multiview stereo). To reduce discomfort, the system included a gaze tracker to find the user's fixation point on an output image and would then calculate the distance infront and behind the point to be displayed in full resolution. The methods of [21] and [22] were used to defocus the other layers.

In addition, when applying the methods of [21] into a system as in [25], the acquired camera images would have already been in focus at a certain level by the camera operator (ex. recall Fig. 1.4.1). Once these images are obtained, there is no way to adjust the depth of focus to sharpen layers that were out of focus during the capturing process. For example, the soccer player slightly to the left of the center of Fig. 1.4.1 appears blurred in the captured image (which would be its maximum resolution). Therefore, the level of sharpness in the original image cannot be increased. Moreover, the method did not perform any subjective evaluation of the resulting natural images.

Before being able to perform alternative smoothing methods for the depth of field replication, further investigation of the human perception of smoothing stereoscopic multiview images needs to be undertaken.

1.4.2 General Spatial Filtering of Multiview Stereoscopic Images

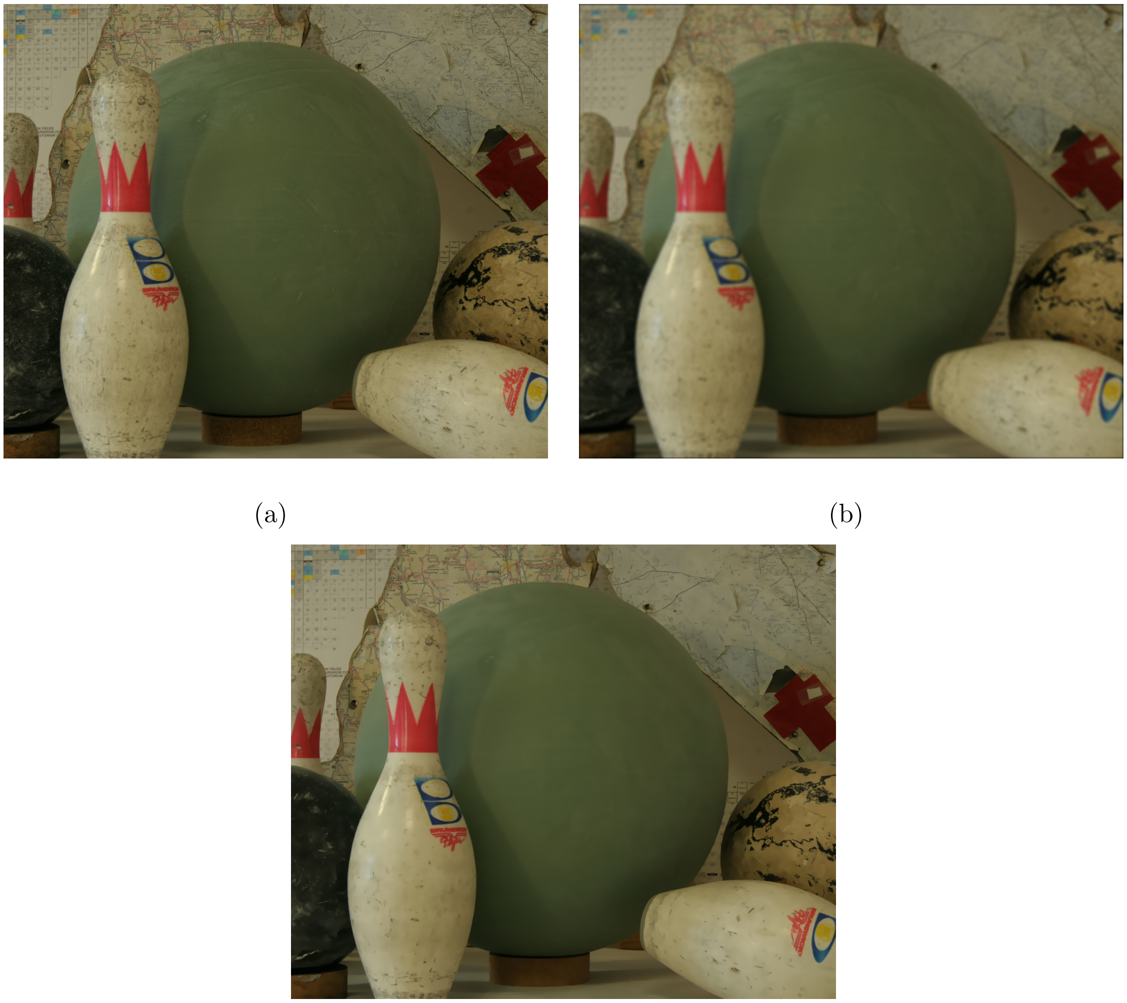
A simple approach to the problem of blurring (smoothing) a stereoscopic image would be to use a two-dimensional (2D) spatial filter and apply it separately to both the left and right images. This would achieve considerable smoothing, however, as mentioned earlier, this method would not respect the depth of the image and it would lead to blurring of the foreground and background components since the filter does not consider depth (see Fig. 1.3(a), 1.3(b)).

Using a 2D bilateral filter would allow for edge-preserved smoothing (see Fig. 1.3(c)), however it too would require the filter to be applied to each view separately which can be time consuming when required for a large set of images such as in multiview video. Using the 3D bilateral filter could solve this problem, but it would require intricate meshes of data (more often used in computer graphics), which are not available for images acquired by a set of cameras without additional post-processing. Furthermore, the 3D bilateral filter would not be an efficient choice for systems conscious of computation cost, where image quality does not need to be perfect (i.e. cell phones, mp3 players, etc.) but instead, would prefer to have quick computation time. Which leads to an unanswered question: in general, how can one perform spatial filtering of stereoscopic multiview images?

1.5 Thesis and Motivation

To date, no method has investigated, or formally addressed how to successfully filter a multiview stereoscopic image using the LDI representation from a general image processing perspective for image smoothing and sharpening. The benefits of developing a general case stereoscopic spatial filter would be the application of filtering on various types of images (both synthetic and natural), in addition to having a method that generates natural looking stereoscopic images. The major challenge to incorporate filtering in a stereoscopic context is how to effectively consider the depth of each pixel when computing the filter output and understanding what impact spatial filtering has on the human perception (viewer discomfort and naturalness of the images) of the resulting image.

The benefit of developing a filtering technique for multiview images would allow post-processing systems to have the ability to perform realistic spatial filtering, and generate



(a)

(b)

(c)

Figure 1.3: Image filtering without considering depth. (a) Original image of the Middlebury stereo dataset ‘Bowling2’ as mentioned in [2]. (b) Result of filtering (a) using a 9×9 averaging filter. (c) Result of filtering (a) using a 2D bilateral filter of size 9×9, $\sigma_r = 4$ and $\sigma_s = 4$.

natural looking stereoscopic images. However, another major challenge stands in the way of incorporating natural filtering on multiview stereoscopic systems such as those using more advanced image representations as layered depth video. Specifically, one challenge that exists in creating a filtering technique for this representation would be to account for the sparse distribution of pixels residing in occluded layers of the image.

To solve these problems, this thesis presents a 3D adaptive filtering technique, and an investigation of its impact on human stereoscopic 3D perception based on the experimental data collected from a group of humans. During the experiment, the human subjects assessed both the naturalness and viewing discomfort of stereoscopic multiview images filtered using the well-established 2D bilateral filter and the proposed 3D adaptive filter. The results will show whether or not either of them can produce a more natural filtering result, which can lead to lower discomfort, what optimal filter coefficients are, and how much quality is given up for increased processing time. These results will become useful in future research/imaging systems, since they will aid in developing a better understanding of how to design a set of filters to be used on already acquired multiview image sets or multiview video systems.

The results will also be useful in the future by aiding in the design of other complex imaging systems, such as an alternative to the ones that can blur multiview images to mimic the human visual system (with the aim of reducing viewer discomfort), or for use as a building block in pre/post-processing steps for compression, noise removal, edge-detection, or image enhancement. The proposed method can also be easily modified (by adjusting filter coefficients) for extending other 2D image operators to 3D for future research purposes or validation.

1.6 Outline

The remaining chapters of this thesis will contain the material outlined below.

Chapter 2: Background Information on Stereoscopic Imaging

This chapter describes background information on stereoscopic imaging that is intended for a reader who is unfamiliar with the domain of stereoscopic image processing. It contains information related to how humans perceive stereoscopic 3D, and gives a brief outline of some of the historical events that have led up to the current state of knowledge. Various forms of technology (past, present and future) used to display stereoscopic images are described.

Chapter 3: Layered Depth Images and Rendering

This chapter describes the fundamentals regarding the image representation of Layered Depth Images (LDIs). Topics that are covered within this chapter include how LDIs are generated, and how they can be rendered. Additionally, the chapter also covers the implemented techniques used to improve the rendering capabilities of already existing methods.

Chapter 4: 2D Image Filtering

This chapter covers the basics of linear spatial filtering and describes some of the common filters used for image smoothing and sharpening. The chapter also covers the more advanced 2D bilateral filter that can be used for edge preserving smoothing.

Chapter 5: 3D Spatial Filtering for Multiview Images

This chapter covers the proposed method for filtering multiview images using the LDI representation. The mechanics of the filter, as well as the algorithm used to adaptively change the filter size to tailor it to the varying size of the depth layers in the LDI are explained. As well, the chapter also describes the experiments performed on human subjects to evaluate the capabilities of the proposed filtering method against using the 2D bilateral filter. Filters for stereoscopic (3D) smoothing and sharpening are proposed.

Chapter 6: Results and Evaluation

This chapter presents the results obtained from using the proposed filtering method on typical stereoscopic multiview images. Some of the images used in the experimental study are also presented, in addition to the subjective results obtained from the subjects. The results of the proposed filtering method are compared to the 2D filtering techniques and analyzed in terms of visual, computational cost, and the evaluation of their usefulness for differing applications.

Chapter 7: Conclusion

This chapter presents a summary of the thesis, the scientific contribution and future work to be investigated.

Chapter 2

Background Information on Stereoscopic Imaging

Imaging systems have become second nature in our everyday lives. From televisions, to billboards, to the little screens on our portable music players and cellular devices. All of these screens are two-dimensional and they can be adequately perceived with one eye. In contrast to the actual world that we see, which requires two eyes to distinguish depth, these 2D screens lack the idea of true depth and rely on monocular depth cues to project a realistic image onto a 2D screen.

To produce images that make use of both of our eyes, the process of *Stereoscopic Imaging* can be used. Stereoscopic Imaging can be described as any technique capable of recording three-dimensional visual information or creating the illusion of depth in an image.

2.1 How Humans Perceive 3D from Two Separate (2D) Images

As humans, we are born with two eyes that see the world from slightly different positions. It is through these slightly different views (*horizontal disparity*) of the same scene that our brains are able to perceive depth in a process known as *stereopsis*. Therefore, if a human is shown two-dimensional images of the same scene taken from different viewpoints (similar to the displacement of the eyes), where each eye is shown only a particular view, then the brain can fuse a 3D image with depth. This technique is used in stereoscopic imaging to produce stereoscopic 3D images.

In 2D imaging, each eye is shown the same image. Since each eye does not see a slightly different view, the image will appear on the screen and have no depth. On the other hand, if the images for eye each are shown crossed-over (negative parallax), the viewer will see an image appearing infront of the screen, thereby creating an image with depth. Conversely, if the two images are separated by a distance that is less than the inter-occular distance (i.e. the distance between the two eyes), than the image will appear behind the screen or what is known as positive parallax. These notions of perceiving 3D are shown in Fig. 2.1..

2.2 History of Stereoscopic Imaging

The following section describes some of the events and technologies that have developed single view stereoscopic 3D into what it is today.

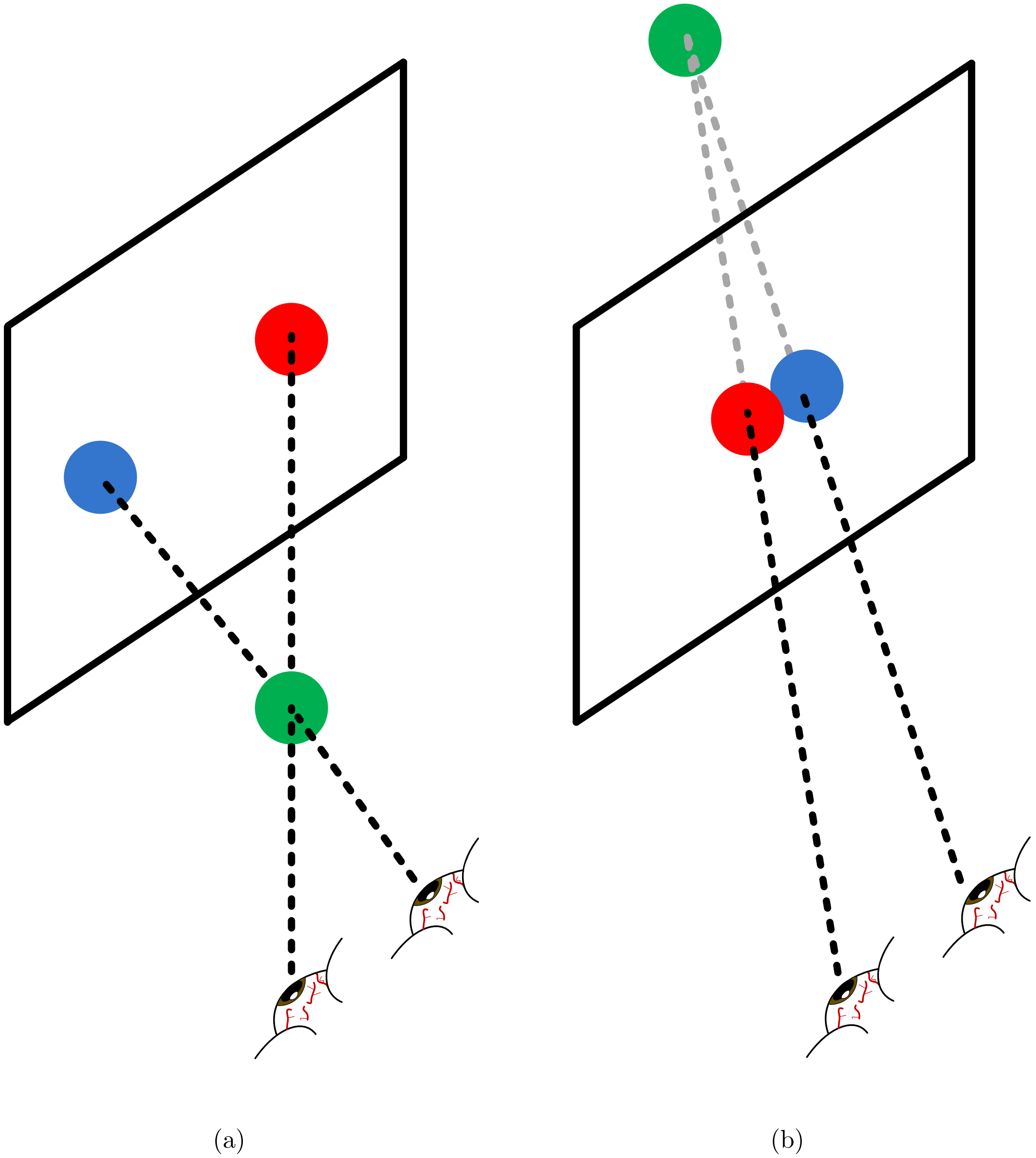


Figure 2.1: Depth from two separate images. (a) Eyes are crossed (negative parallax), the image appears in front of the screen. (b) Example of positive parallax where the image appears behind the screen.

2.2.1 First References - Euclid, Artistic Effects of Monocular Depth

Some of the first references to an interest in binocular vision can be attributed to the mathematician Euclid (300 BC). He considered the problem of observing a sphere with two eyes and noted that if a viewer directly faces a sphere, then using one eye, they will see a slightly smaller sphere (likewise with only using the other eye). He also mentioned that using both eyes together, the viewer would be able to see half the sphere if its diameter was equal to the distance separating both eyes [26].

Knowing that depth can be perceived using a single eye, artists were able to convey information about the depth of a scene inside a 2D painting making it appear more realistic. Using these forms of monocular depth cues, artists were able to make a 2D image appear to be 3D. Some of these techniques include converging parallel lines, texture gradients, and relative size of objects. Texture gradients usually reveal objects that are closer to the viewer will have more small detail information, while objects that are further away will have less detail. In terms of size, objects that are larger in a scene are typically closer to the viewer, while objects that appear smaller are further away. Lastly, converging parallel lines usually indicate that the closer they are to each other, the further away from the viewer. These ideas were shown in Fig. 1.1(a). Many of these ideas are also present in 3D computer graphics when rendering 3D geometry onto a 2D screen. Similarly, a set of shaders, and projection geometry is used to to make a 2D images appear as though it were 3D.

2.2.2 Stereoscope and First Stereoscopic Movies

The first time that an individual was able to reproduce a stereoscopic image using two (separate) 2D images was during the 1830s when Sir Charles Wheatstone invented the "Stereoscope" [26]. The Stereoscope was a device where each eye of a viewer could see a distinct 2D image using a mirror (shown in Fig. 2.2). Sir Wheatstone was later credited with the publication of the discovery of stereopsis, the process by which the brain can perceive depth by looking at two (slightly different) 2D images of the same scene.

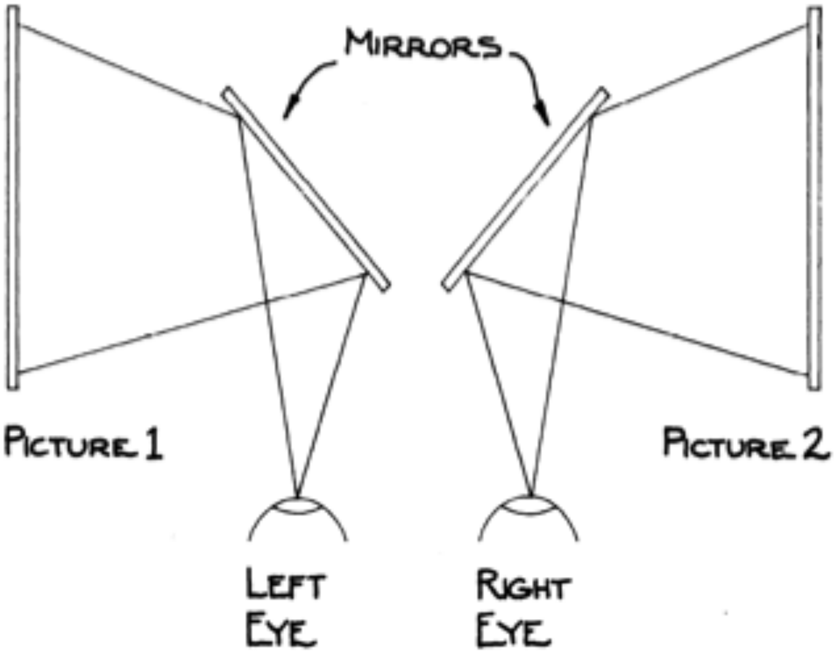


Figure 2.2: Sir Charles Wheatstone's Stereoscope. Image was obtained directly from [26].

2.2.3 Anaglyph Technology

Since the 1830s, anaglyph technology has been used as a method to separate left and right views to facilitate the process of stereopsis. The anaglyph method uses complementary colour filters to separate the view for each eye. An example anaglyph image is shown in Fig. 2.3.

The anaglyph method of vieweing stereoscopic images is an inexpensive method because the filter glasses are inexpensive, and the production challenge only lies in separating the left and right views by using different colours. Therefore, it is easy to produce low quality stereoscopic 3D for a large audience. The main disadvantage of this system is that it has poor colour reproduction.

Some of the most common uses of stereoscopic anaglyph has been seen in old Hollywood movies, and also in comic books. This technology is still present in computer games (the *Nvidia 3D Vision Kit* has an anaglyph option), and it is also still used to produce

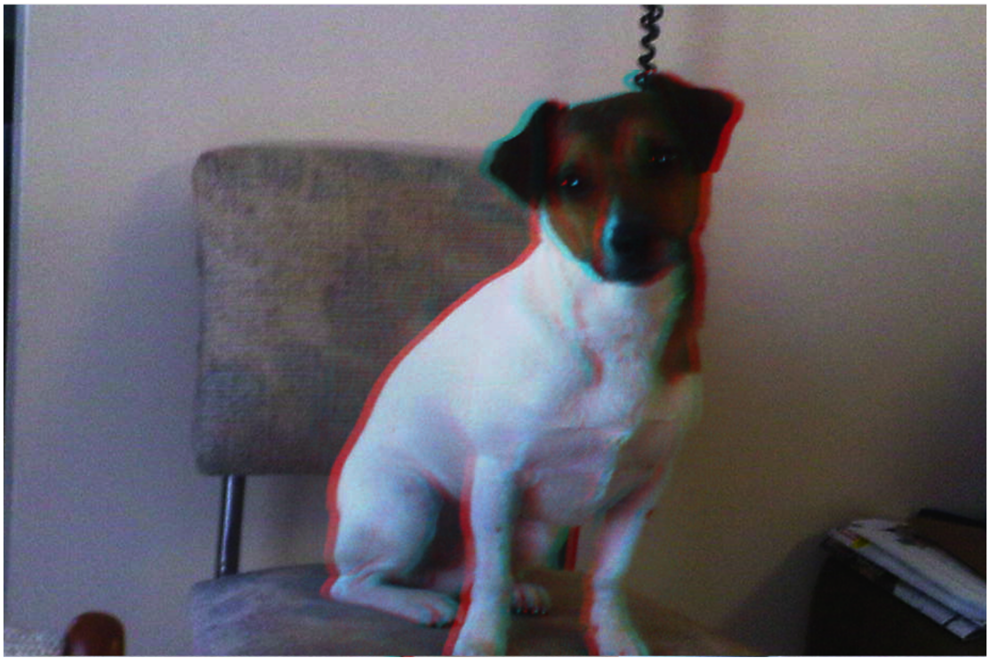


Figure 2.3: Anaglyph image of a dog.

stereoscopic television programs using standard definition televisions or high definition televisions (along with the 2D broadcasting backbone). Some of the recent events shown on TV in Canada using anaglyph stereo were the, “Queen Elizabeth in 3D,” shown on CBC on September 20, 2010, which was also the first 3D documentary broadcast in Canada [27], and MTV Live shown on MTV Canada on March 17, 2011 [28].

2.2.4 Recent Technology and Applications

The stereoscopic technology that is widely being used today still has integral components that date back many decades. We will limit the discussion in the following sections to passive stereo, active stereo, and autostereoscopic technology that is intended for single view stereo. A discussion of multiview stereoscopic technology that currently exists, and that is envisioned for the future, will be briefly discussed.

Passive Stereo

Passive stereo was first used in the 1930s with the aid of the polarization of light. Using the phenomenon of polarized light, left and right images could be projected using circularly polarized light and could then be selected using glasses with opposite polarizing filters on each eye. This is another cost effective method for theatres since the glasses are inexpensive and only a projector is needed that can project two images using orthogonal polarizations of light (and a silver screen - i.e. one that maintains the polarization of light as best as possible).

Passive stereo is most often used presently in 3D cinema (ex. IMAX 3D and RealD 3D) since the glasses are cost effective and images produced are high quality with no loss of colour. The problem that exists with this method is increased “ghosting” that results when one filter cannot fully reject the opposite signal. This is usually caused since the projectors and screen do not fully conserve the polarization of light causing the image to leak into the wrong eye.

Another use of passive stereo that is slowly picking up momentum is present in 3D televisions. Manufacturers such as JVC [29], and LG [30], are using passive stereo technology since consumers can have high resolution 3D, without having to pay upwards of $100 for each set of glasses (which can become expensive for a family of four or more). Using similar glasses to the ones in the theatre, the cost is shifted toward the actual television set. To produce polarized light, the television is setup with a polarizing sheet on the outer layer that makes opposite polarizations for every other row of pixels. This results in a lower screen resolution, but it is cheaper in terms of glasses, and it can still accurately reproduce colours. However, much like the theatre, it can nonetheless be hindered by ghosting effects.

Active Stereo

Active stereo technology uses the concept of “active” glasses. These glasses typically have LCD shutter lenses that rapidly turn off (or block) the image from one eye by not allowing any light to pass through. During this time, the first image of the stereo pair is shown on the screen (in full resolution) and is viewed in the first eye. Subsequently, when the source shows the following frame in the sequence, the lens that was previously open becomes opaque, and the previously opaque lens becomes transparent. This allows the other eye to view the second image of the stereo pair. The frame rate used is typically 120 frames per second so that the brain cannot distinguish any lag in the image sequence, but rather perceives a continuous video stream from each eye.

The increase frame rates are a requirement that must be met for active stereo. In the past, movie theatres used active stereo using a viewport in front of each chair in the cinema [26]. These viewports were mechanically controlled by a motor that alternated with the images in the film. However, due to lack of synchronization and restricted positioning, this method was abandoned. More recently, active stereo has been realizable using glasses which are triggered using a signal that is synchronized from the video source (either wired or wirelessly using infra red). These glasses are typically very expensive, but the display source does not have to giveaway any screen resolution in the process (such as passive stereo TV that gives up every other row for each eye). Active stereo is usually the choice for applications requiring the greatest level of visual quality.

The typical applications of active stereo used presently include 3DTV (the ones with the “cool” looking glasses), or in computer applications for visualization of 3D images or games. One example of such a system is the *Nvidia 3D Vision Kit*. These computer applications require a special monitor that has a refresh rate of 120 Hz and a graphics card equipped with specific Nvidia GPUs.

Autostereo

Up to this point, all of the technologies used to display stereoscopic 3D required the use of special glasses to separate views for each eye. In contrast to these types of technology, an autostereoscopic 3D system is one that does not require the use of special glasses to separate views. Autostereo typical use a special cover that is made up of a 2D array of tiny lenticular lenses above each pixel of the original source. The purpose of the array of lenses is to direct light (from a pixel) toward an appropriate eye. Please see Fig. 2.4 for an image referring to this process.

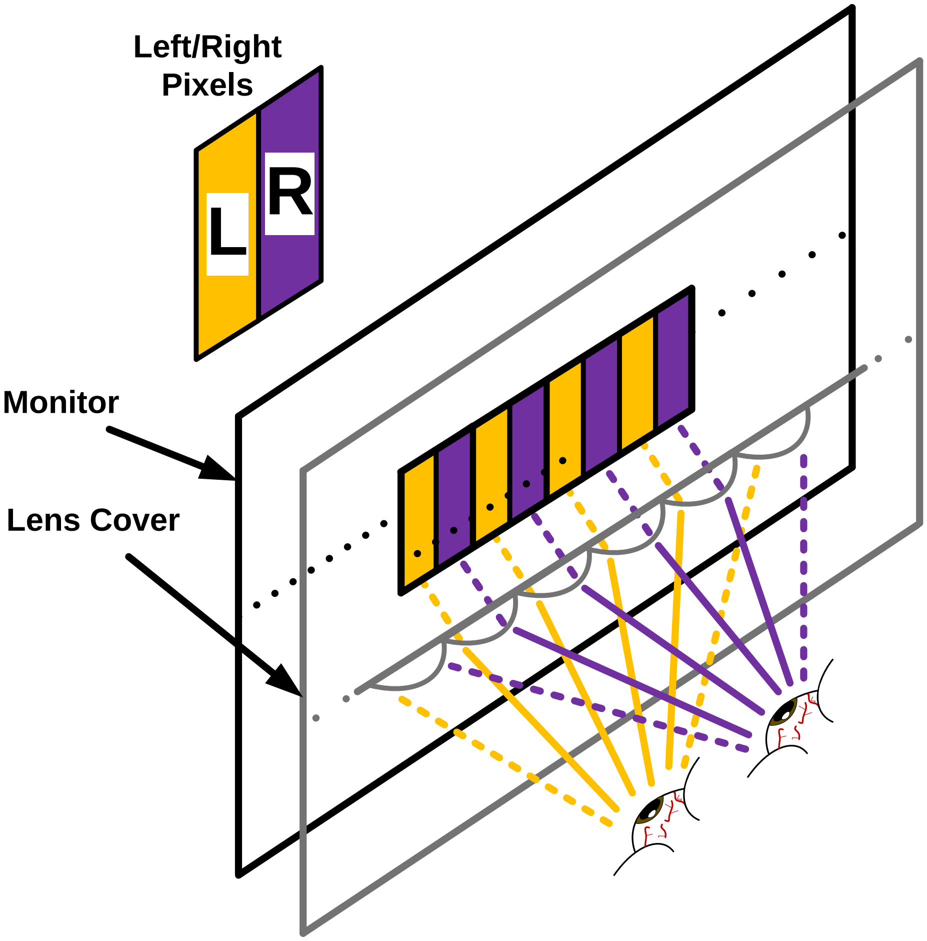


Figure 2.4: Autostereoscopic Monitor Example with Lenticular Lenses.

Autostereoscopic technology is only possible by reducing the spatial resolution of the display source. Consequently, a display source that is setup to accommodate many viewers in distinct “channels” will result in a decrease in available resolution for each user. In addition, each viewer will be confined to what is known as a, “sweet spot”, where the user must have their head in order to see the 3D effect, otherwise, they will see a double image. This poses a restriction on the position from where one is allowed to view the display.

Currently, autostereoscopic televisions exist, but are more expensive than active/passive stereo televisions. However, the advantage of not having to wear a set of glasses has the industry gearing towards this technology as the way for the future [31]. Eye-tracking systems are in existence, with this technology the position of the viewer’s head is tracked and fedback to the display where the picture is adjusted to meet the changing position.

The applications where autostereoscopic displays are most commonly found today, besides televisions, are in handheld devises, or other systems that have a good idea of the location of the viewer. For example, 3D screens for the iPhone, the Nintendo 3DS, Fuji FinePix 3D digital camera, all use autostereoscopic screens because the position of viewer will not deviate too much from the normal orthogonal view (which is directly in front of the screen). This idea is also useful in laptops or computer monitors where the viewer typically has a narrow viewing angle. In contrast, viewers of autostereo television sets may wish to view the screen from a wider viewing angle and thus, it is more difficult to predict and accommodate. This may pose problems for buyers who wish to align the autostereo television with the seating arrangement in their living rooms.

2.3 Future Outlook of Stereoscopic 3D Applications and Research

Having a separate image for each eye helps to re-create a more life-like image or video, after all, we have two eyes. Likewise, most display systems will eventually become stereoscopic once a few of the kinks are worked out, some of which are currently plaguing the success of 3DTV in the consumer market.

Some of the future applications that will be focused on stereoscopic technology will be teleconferencing/telepresence, the use of stereo imaging for machine vision, broadcasting, television, handheld/mobile devices, advertising, and medical.

2.3.1 Medical Applications

The are many applications that can be used for stereoscopic displays in the medical field. There have already been accounts of doctors using computer vision during surgery to make gestures to a camera with depth to cycle through pictures [32]. Likewise, the power of stereoscopic images will allow for a better reconstruction of images taken where a medical specialist will be able to see a 3D image, which is what our body actually is (since our bodies are not 2D structures).

Medical applications will also benefit from stereoscopic technology since viewing stereo is linked with higher decision making confidence [33]. It would therefore be possible to use stereoscopic images to help doctors make better decisions when viewing images of the body in 3D (rather than 2D).

2.3.2 Consumer Applications

Since 2009, the consumer has been bombarded with hype regarding 3DTV. Two consecutive consumer electronics shows (2010 [34] and 2011 [35]) have been centered around television manufacturers and the release of their latest televisions, some with glasses, and others without. They are all pushing for the consumer to get onboard, but that has yet to be adopted with open arms.

Many consumers have just recently made the jump to High-Definition. Most often, the concerns that are holding back the average consumer from diving headfirst into this new form of technology include, the increased cost, lack of content, and uncertainty with the technology (i.e. will this be a fad?).

The questions involving content are being addressed in combination with industry and research. Many companies are releasing 2D to 3D conversion systems so that archived material can be re-released in 3D format to help generated new revenue streams for different markets.

But, there are other problems which can be classified as more important in the eyes of the consumer. These problems are those involving all aspects of viewing discomfort, having to wear glasses, and not being impressed with the images they are seeing. This usually happens when a viewer seeing 3D for the first time, sees a poor 3D image, leaving them dissatisfied. However, with each of these problems, there exists tremendous potential for future research and growth to occur. One area that can lead to more impressive 3D is the concept of multiview stereo that was introduced in the first chapter.

Multiview

As mentioned in the introduction, multiview stereo will be the next form of 3D since it will offer the viewer a more immersible experience by allowing them to view the scene from any position. This will enable more life-like telepresence and make watching tv and movies a lot more of an ‘active’ experience for the user (not simply a passive system). More details regarding the viability of multiview stereo for the future of stereoscopic imaging can be found in the introductory chapter (Section 1.2).

Volumetric Displays

As it was briefly mentioned in the introduction, another future technology that will improve the stereoscopic viewing experience are volumetric displays. These displays are considered to be the optimal stereoscopic display systems of in the future. The reason that volumetric systems will be successful is because the viewer will not require any glasses, and since the structure will appear as a volume in space, it will not lead to the vergence-accomodation mismatch that happens when viewing stereoscopic images on a 2D screen.

Researchers are currently working towards making volumetric displays achieveable in the near future. Specifically, a current limitation exists in the refresh rates. One group has already developed a system that is capable of two-second refresh rates, which is promising since only a few years ago the highest refresh rates were on the order of minutes [1]. In addition, this same group has a prototype of a 10-inch screen and is looking towards increasing display size, improving resolution, and reducing system size and power usage.

Chapter 3

Layered Depth Images and Rendering

Layered Depth Images are commonly used in computer graphics, but also for multiview imaging since they can be used efficiently for the storage and transmission of image data. Only the main layer, plus residual data and depth are needed to be stored/transmitted, while the receiver can then render any view of the scene similar to depth-image based rendering.

3.1 Background on Layered Depth Images

A Layered Depth Image (LDI) as proposed in [10], is a representation for a three-dimensional scene using an array of two-dimensional image slices (also known as layers or an array of layered depth pixels as defined in [10]). Each LDI has a reference view point from which only those pixels in the first layer (i.e. the pixels that are seen as the closest to the reference view) are visible while the rest of the pixels in other layers are occluded from this view as shown in Fig. 3.1. The benefit of using this representation arises when one wishes to render a 2D view of the three-dimensional scene as seen from a different near-by viewpoint rather than the reference view since the pixel information from layers

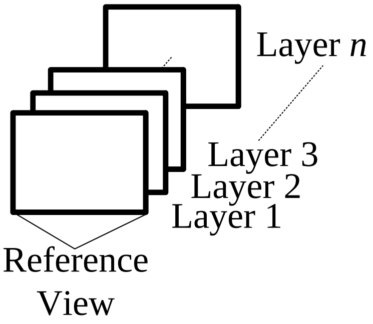


Figure 3.1: Layered Depth Structure.

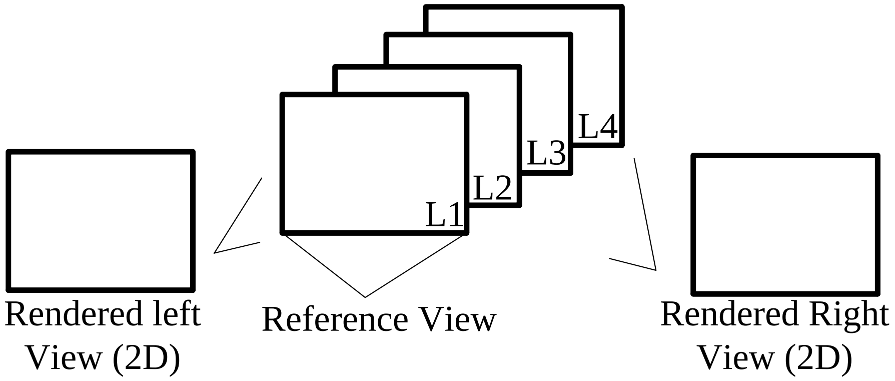


Figure 3.2: Rendering different views.

2 to n can now become visible (if they are seen as closer to the viewer than pixels in layer 1) as shown in Fig. 3.2. These newly visible pixels can then fill in regions that would have otherwise been occluded if only a regular (i.e. not an LDI) image was warped to the nearby viewpoint. This is particularly useful in multiview video and free viewpoint video since the LDI can be used to interpolate high-quality views between known camera views [5]. A similar argument regarding the practicality of the LDI can be made in the context of stereoscopic imaging since the representation can render arbitrary left and right eye (2D) images from a 3D scene where the additional layers in the LDI can be used to fill in disocclusions at different viewpoints.

3.2 How to Generate Layered Depth Images

Layered depth images can be generated using either acquired camera images and depth maps, real images taken from 360 degrees around an object, or computer graphics images. Each of these methods of generating an LDI are described in the subsections below.

3.2.1 Acquired Camera Images with Depth Information

An LDI of a 3D scene can be constructed using a set of 2D views taken from different viewpoints and the depth map of each of these viewpoints as shown in Fig. 3.3. for a four view case. The depth maps can be estimated in a number of ways including: i) computing pixel correspondences across images [36], [37], ii) using classification [38], [39], iii) analysis of object motion [40], [41], iv) human assisted techniques [42], or v) structure light [43]. In Fig. 3.3, since the four camera views capture the 3D scene from different viewpoints, they require warping into a common reference view to construct the LDI. This can be achieved by warping each of the views in Fig. 3.3 similar to (3.1). Once all views and depth maps are warped to the reference view, the LDI can then be constructed by sorting the pixels with the same (x, y) coordinates according to their depth values. For example, the warped versions of each camera view will have different depth values for its respective pixels. Of these warped pixels, the ones that are seen as closest to the reference view (i.e. pixels that have the smallest depth) will be used to construct the first layer of the LDI. The remainder of the pixels are then categorized by depth value, and sorted into layers 2 to 4, where layer 4 is the furthest away. When the difference of the depth values between pixels (with identical coordinates) in adjacent layers are less than a preset epsilon (as in [10]), the average pixel value is placed in the closer layer (to the viewer), while the other layer is left empty.

An example of warping four of the camera views from one frame of the breakdancing set [5] is shown in Fig. 3.4. For simplicity, in Fig. 3.4, the reference view was selected

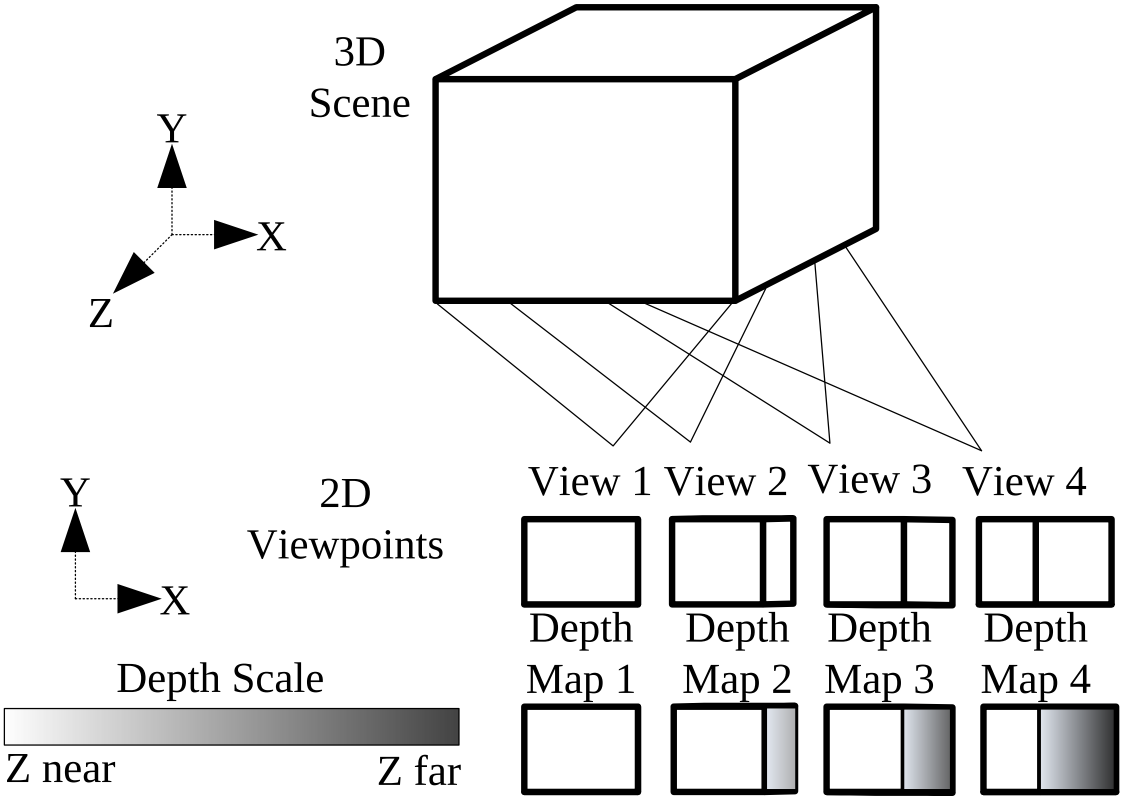


Figure 3.3: Multiple camera views and depth maps from a 3D scene.

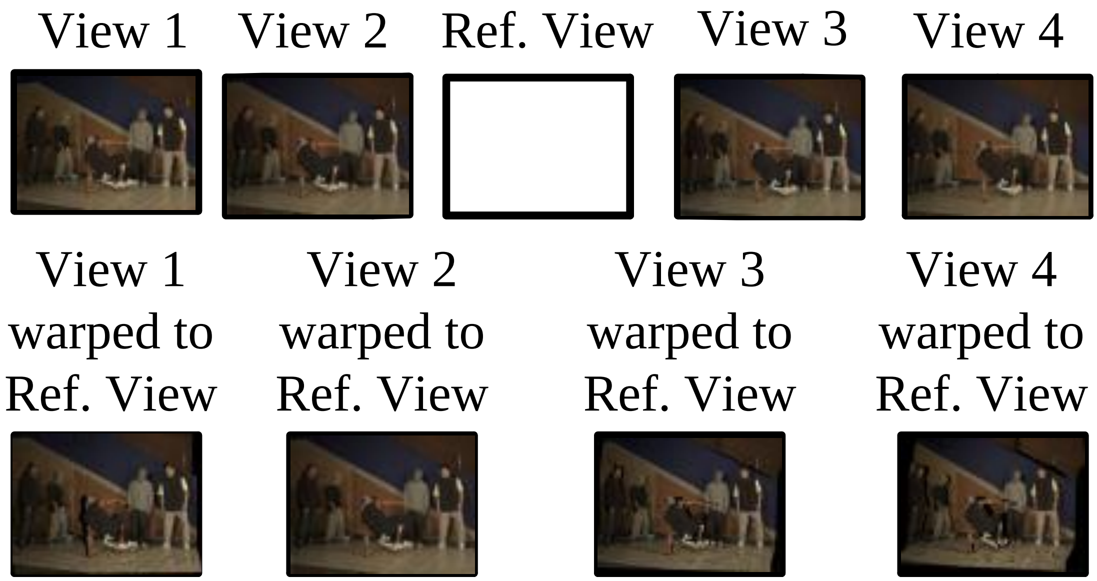


Figure 3.4: Warping views to the reference view.

to be that of view 2 (to reduce the number of times a view needs to be re-sampled). As shown, when warping view 2 to itself there is no change.

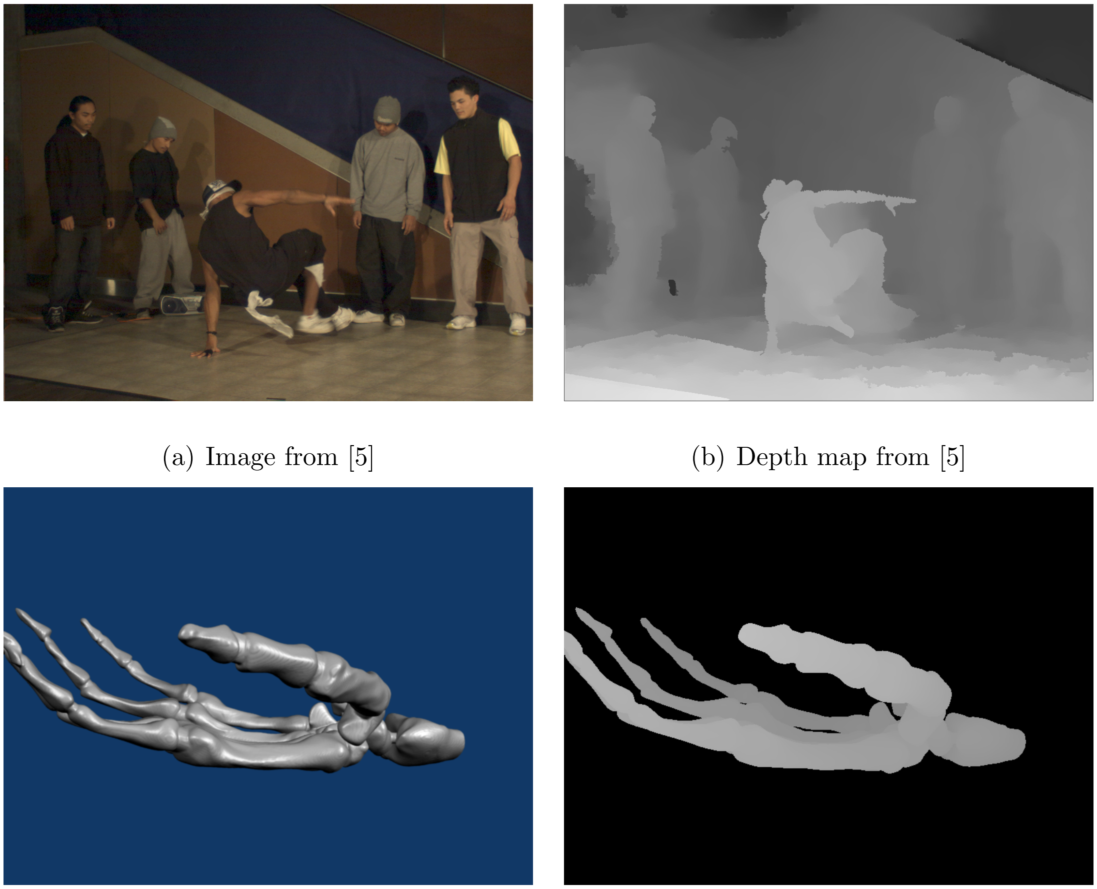
3.2.2 Computer Graphics Models

The LDI representation can also be used for computer graphics images. Normally, computer graphics images are represented as intricate meshes of data in a 3D space. To generate 2D views of objects to be presented on the screen, the 3D mesh is projected onto the 2D view in a process known as rendering. The difference with the LDI representation is that less data is used to store the information presently in the 3D scene. Only the pixels that are visible in the reference view are stored in the first layer, followed by any subsequent pixels that have additional depth information. The depth information of the scene can then be obtained using the *z-buffer*.

The benefit of using computer graphics images is that the depth information returned from a *z-buffer* is more accurate than the depth maps obtained through the computation of pixel correspondences or depth map estimation (which is used in the case of acquired camera images). A comparison of depth maps generated from these previously mentioned methods is shown in Fig. 3.5.

3.2.3 Real Images from 360 Degree Rotation

LDIs can also be constructed by rotating an object (usually on a computer-controlled turntable) 360 degrees and taking photographs at constant intervals. Using the real photographs of the object from 360 degrees, the LDI can be constructed using a process similar to voxelization (a process where pixels are projected onto candidate voxels similar to [44]), except this time the voxelization becomes view centered similar to the LDI structure.



(a) Image from [5]

(b) Depth map from [5]

(c) A computer graphics image

(d) Depth map obtained from (c)

Figure 3.5: Comparison of depth maps generated from pixel correspondences (acquired camera images) and depth maps generated from a z-buffer (computer graphics). The model for the "hand" in (c) was obtained from the *Stereolithography Archive* at Clemson University.

3.3 How to Render Layered Depth Images

The process of obtaining the coordinates of a point (x_2, y_2) in an output view using the point (x_1, y_1) from the reference LDI view can be described (as in [10]) by:

$$T_{1,2} \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = \begin{bmatrix} x_2 \cdot w_2 \\ y_2 \cdot w_2 \\ z_2 \cdot w_2 \\ w_2 \end{bmatrix}, \tag{3.1}$$

where $T_{1,2}$ is the transfer matrix defined as $T_{1,2} = C_2 \cdot C_1^{-1}$ where C_1 and C_2 are the 4×4 camera matrices of the LDI and output camera, respectively. The coordinates of ($x2$, $y2$) in the output camera view can then be obtained by dividing by w_2.

3.3.1 Problems After Rendering

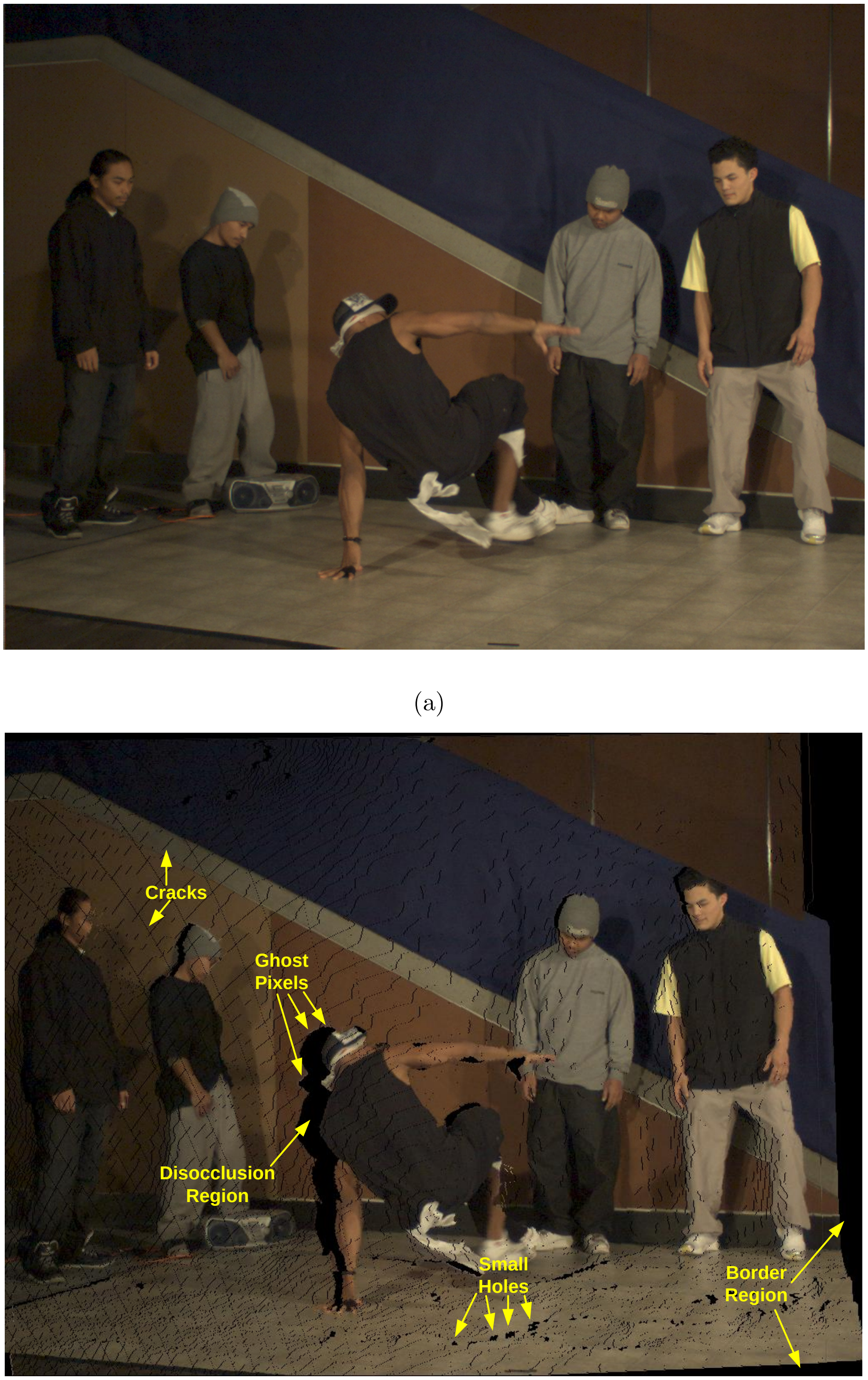
The rendered left and right views should not be viewed immediately after warping since they suffer from: i) Pixel Discontinuities (i.e. cracks in the image), ii) Image Artifacts (or holes), iii) Ghost Pixels, and iv) Disocclusion regions visible. Each of these problems are described below and can be seen in Fig. 3.6.

Pixel Discontinuities

Pixel Discontinuities (sometimes referred to as empty pixels, or image cracks [45]) arise from rounding to the nearest integer pixel coordinate position when warping image views. As a result, certain pixels are left empty and the image appears to have one-pixel wide cracks (also refer to Fig. 3.4 bottom row view 1, 3 and 4).

Ghost Pixels

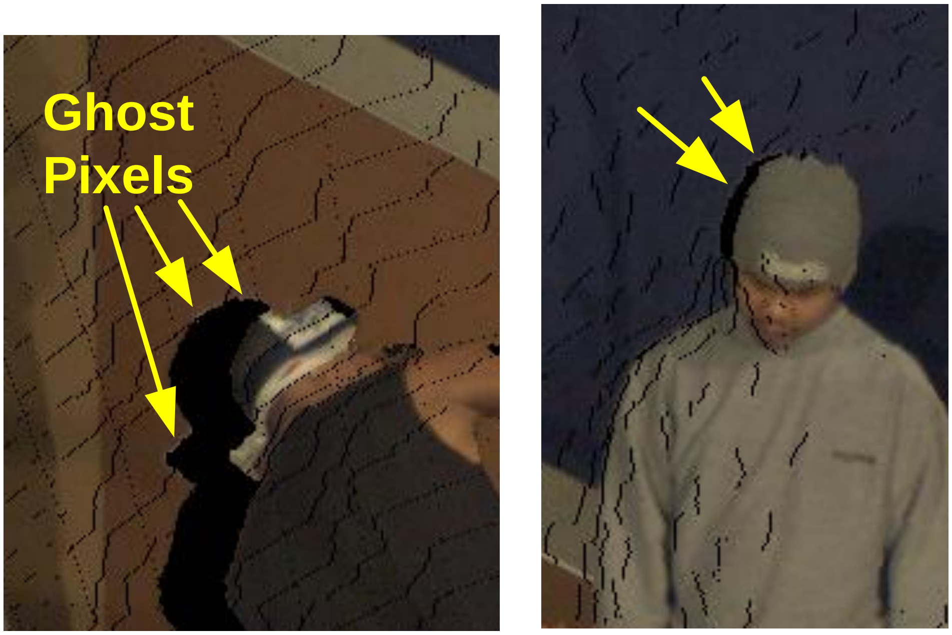
Visible Ghost Pixels is a problem that occurs due to the inaccuracies in the depth map, particularly at discontinuities occurring at object boundaries. The points at the object boundaries contain colour information from both the foreground and background. When warped to the new view, the inaccuracies in the depth map cause these points to be warped to wrong locations causing what is known as “ghost contours”. To aid in the viewing of Fig. 3.6(b), the areas of the image showing ghost contours have been enlarged and displayed in Fig. 3.7.



(a)

(b)

Figure 3.6: Problems after image warping. (a) Original Camera View. (b) Image (a) warped to another view with rendering problems indicated.



(a) (b)

Figure 3.7: Ghost contours after rendering Fig. 3.6(a). (a), (b) sections of the warped image zoomed in.

Image Artifacts and Disocclusion Filling

Image artifacts (also called image holes) are disocclusion regions of pixels that were occluded in the reference view [46]. When warped to the new view, these occluded pixels become visible and appear as a set of holes (since no pixel information is available). A common cause for this problem is due to the sharp changes in the depth maps of the image.

3.3.2 Existing Techniques and Proposed Refinements

To remove these problems, various techniques have been adopted from the papers written on how to improve the rendering of LDIs [45], [46], [47], [48]. These techniques have been adopted to construct a rendering algorithm (shown in Fig. 3.15) to refine the rendered views (and stereo image pairs) from the LDI data. The details regarding the processes used in the algorithm are discussed in the sections below.

Removing Pixel Discontinuities

This error can be corrected by replacing the empty pixel with the median filter value of the 3×3 neighborhood [47] around it. However, unlike [47], our method to render the LDI used the median value of the 7×7 neighborhood around it to obtain a higher degree of confidence. In addition, performing median filtering on the entire image can be tedious (due to long processing time), thus, a labeling of empty points is conducted. Following the labeling of each empty pixel, the median filter is applied to each point in the 7×7 neighborhood of the crack location (see Fig. 3.8(b) and the crack removal result in Fig. 3.8(c)).

This refinement requires less processing time than median filtering the entire image, fills in the cracks with suitable replacements, and corrects other pixels that were warped to incorrect locations (i.e. pixels that were not necessarily empty). Depending on the processing time required, this step can be tweaked to reduce processing time by adjusting the median filter width w. Reducing the filter width comes at the expense of crack filling/pixel replacement suitability (in the event that there is a large degree of noise in the small median filter window for a given pixel).

Removing Ghost Pixels

Ghost pixels can be removed using the labelling technique given in [45]:

$$\forall_{x,y} \in S, \left(\sum_{i=-1}^{1} \sum_{j=-1}^{1} D(x+i, y+j) \right) - 9(D(x,y)) > T_d, \tag{3.2}$$

where S is the image, and D is the depth map of the reference camera. T_d is a predefined threshold as stated in [45]. When (3.2) is satisfied, the point (x, y) is labeled as a ghost pixel on the right-hand side. Subsequently, the pixels flagged as ghost pixels are not warped to the new view during rendering. An additional labeling step is then used for the left-hand pixels.

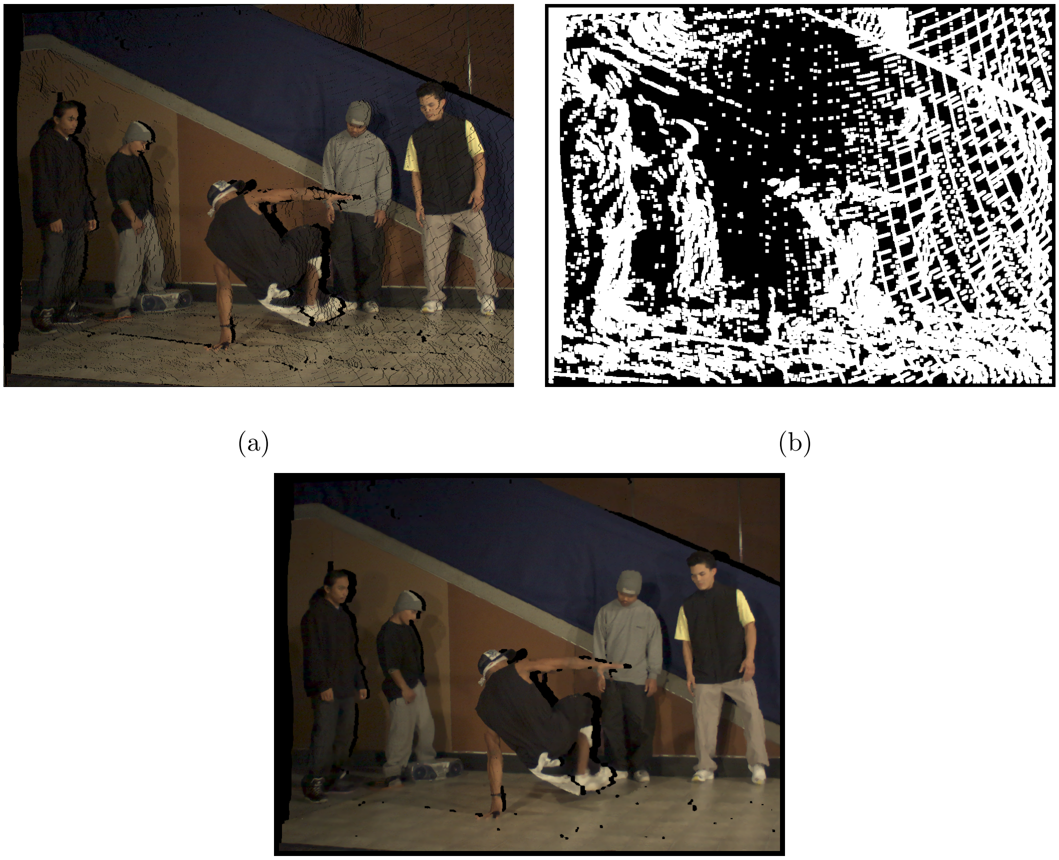

(a) (b) (c)

Figure 3.8: Labeling and removing cracks. (a) Image showing cracks, (b) Labeling of cracked pixels from (a) and their 7 × 7 neighborhood. (c) Result of removing cracked pixels with a 7 × 7 median filter.

A simpler technique to correct ghost pixels, one that has been used throughout image rendering systems to remove outliers from the perimeter of a hole, has been adopted. Namely, the morphological operation of binary dilation of the binary mask BM by SE or $BM \oplus SE$ where the mask BM defines the empty pixel locations with a value of '1' and SE is the structuring element. The dilation operation thereby grows the regions of BM representing the empty pixels (i.e. grows the holes in the image outwards), which in turn removes the outliers which lie on the perimeter of these holes. This operation can be represented by the set of pixel locations z in which the reflected structuring element overlaps with the empty pixel locations in BM or:

$$BM \oplus SE = \{z | (\hat{SE})_z \cap BM \neq \emptyset\}. \tag{3.3}$$

The structuring element used for the purposes of dilating the holes is given in (3.4).

$$\begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \tag{3.4}$$

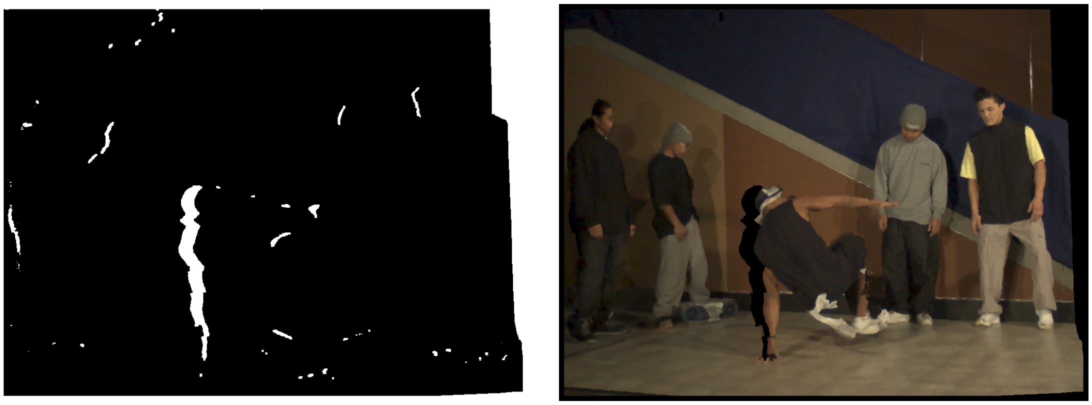
Removing Image Artifacts and Disocclusion Filling

To remove/fill the holes, the rendered image is scanned for the remaining empty pixels and a binary mask M is generated with these candidate hole pixels (see Fig. 3.9(a)). Each of the holes H_i in M are then labeled as either small holes, disocclusion regions, or border holes using:

$$H_i = \begin{cases} \mathit{SmallHole} & \text{pixelcount}(H_i) < T_h \\ \mathit{Disocclusion} & (\text{pixelcount}(H_i) > T_h) \\ & \text{AND}(H_i \subseteq \mathit{Centre}) \\ \mathit{Border} & \text{else} \end{cases}, \tag{3.5}$$

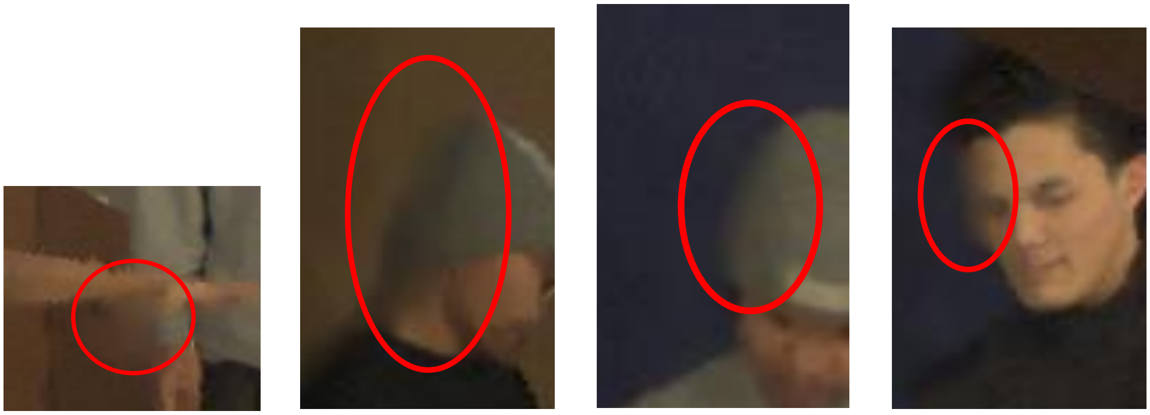
where T_h is a predetermined threshold value, and *Centre* is a mask defining all pixels that are more than 20 pixels away from a border.

The small holes are filled using the average of the non-zero pixels around the boundary of the artifacts as in [48]. The result of filling the small holes in Fig. 3.8(c) is shown in Fig. 3.9(b). This filling method was not used for the large regions because it would result in a unnatural looking image. The method in [48] works well in small regions because the blur is not very noticeable in the background. However, discretion should be used when setting the value of T_h since larger values will result in larger holes appearing blurred. A number of examples indicating the limitation of this filling method are shown in Fig. 3.10. It is therefore preferable to limit the use of this filling method to small holes in the



(a) (b)

Figure 3.9: (a) Binary mask M of empty pixels after crack filling and dilation of holes. (b) Result of filling the small holes with the method of [48].



(a) (b) (c) (d)

Figure 3.10: (a)-(d) Examples of hole filling using interpolation in detailed regions of the image. These images were generated using a $T_h = 200$ pixels.

background of the image since the performance of interpolation is not easily discernible, and instead, use more sophisticated methods for larger holes. The procedure for filling large holes is described next.

The large disocclusion regions in the rendered image are filled by first propagating the background pixels across the empty region, as in [46] using:

$$\begin{array}{lcl} p_{fg} \in \partial\Omega & \rightarrow & p_{bg} \in \partial\Omega \\ B_{\varepsilon}(p_{fg}) & \rightarrow & B_{\varepsilon}(p_{bg}) \end{array}, \tag{3.6}$$

where p_{fg} are foreground points on the boundary of the disocclusion region $\partial\Omega_{fg}$ (within the foreground region). Equation (3.6) implies that these foreground points on the boundary are replaced with the background points p_{bg} on the boundary $\partial\Omega_{bg}$. $B_{\varepsilon}(p_{fg})$, the known neighborhood around point p_{fg}, are replaced with the known neighborhood $B_{\varepsilon}(p_{bg})$ around point p_{bg}. The source pixel can also be extracted at a finite offset from the disocclusion region boundary (see 'Offset Width' in Fig. 3.11). This replacement process takes place across pixels with the same relative y-position, and is repeated for a six-pixel thickness (see 'Replication Width' in Fig. 3.11).

After the manipulation of the image, the remainder of the hole is removed using the method of Criminisi *et al* in [49]. However, for our purpose the source region Φ as defined in [49] should ideally be set to

$$\Phi - p_{fg}, \tag{3.7}$$

which is the source image without the foreground pixels since the disocclusion region typically contains components of the image background.

To compute the appropriate region of $\Phi - p_{fg}$ for each large hole, the first step taken is to determine a crude approximation of a source region to reduce the eventual search area. This crude source region is deduced by determining the maximum and minimum column and row values of the large hole and then extending each by a pre-determined factor λ. The source region can then be represented as a rectangular subimage from the original (which does not exceed the original image boundaries) with corners defined as,

$$\begin{array}{ll} (x_{min} - \lambda, y_{min} - \lambda), & (x_{max} + \lambda, y_{min} - \lambda), \\ (x_{min} - \lambda, y_{max} + \lambda), & (x_{max} + \lambda, y_{max} + \lambda). \end{array} \tag{3.8}$$

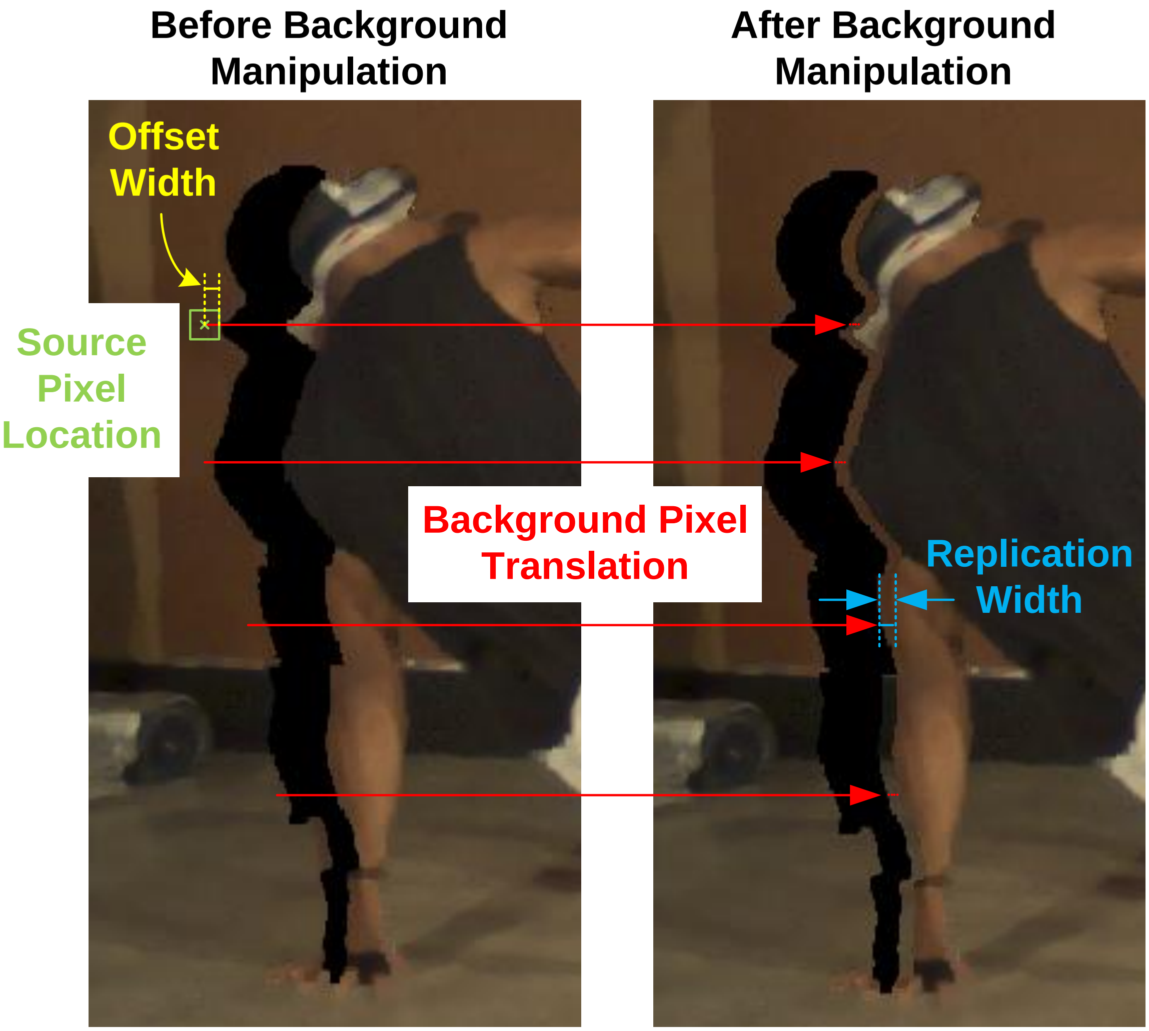
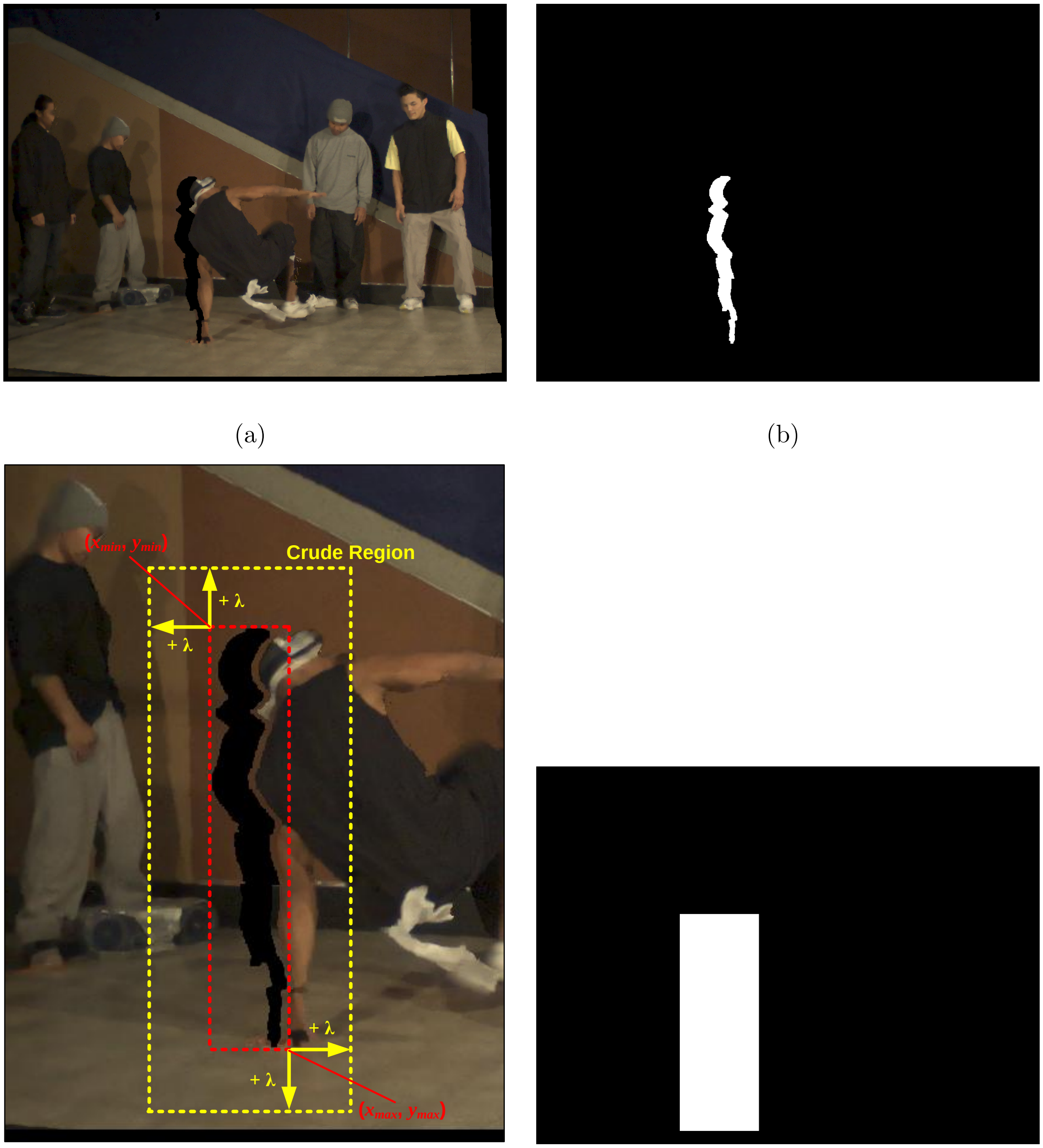

Before Background Manipulation
After Background Manipulation
Offset Width
Source Pixel Location
Background Pixel Translation
Replication Width

Figure 3.11: Example of background manipulation.

Increasing the value of λ will result in a larger source region, which will return a more accurate hole inpainting at the expense of processing time since a larger search area will exist (see an example of a crude source region in Fig. 3.12(d)). Using this value for $\Phi - p_{fg}$ is a crude approximation and will not work well for filling large holes since the crude source region is likely to contain foreground points adjacent to the large hole. Therefore, the foreground points of the image (which can be obtained by thresholding the depth map as in Fig. 3.13(b)) must be removed from the crude region. An example of such a source region without the foreground points or pixels within the hole itself is



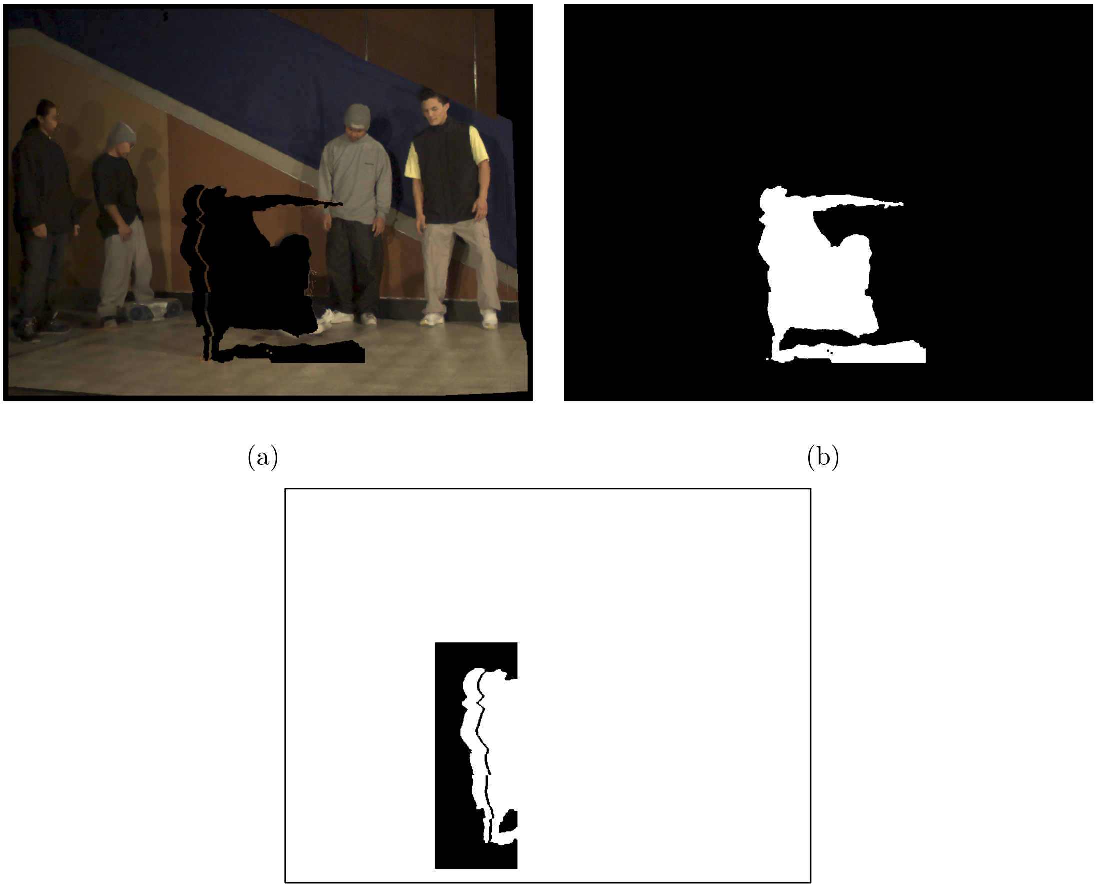
(a) (b)

(c) (d)

Figure 3.12: (a) Image with disocclusion (Fig. 3.9(b) re-shown for convenience). (b) Binary mask indicating large hole location. (c) Crude selection of the source region extended by λ pixels from the extremities of the disocclusion region. (d) Binary mask of crude source region for $\lambda = 50$ pixels.

shown in Fig. 3.13(c).

The result of using this source region to fill the disocclusion region is shown in Fig.



(a)

(b)

(c)

Figure 3.13: (a) Image with disocclusion region and foreground pixels (with depth map intensity range of 55 to 60) removed. (b) Binary mask of the foreground region. (c) Binary mask of the final source region. Pixels with value ‘0’ indicate the final source region.

3.14. The border pixels are also filled using the method in [49], however, this time background pixels are not propagated across the region and Φ is reverted back to its original definition.

3.3.3 Proposed Rendering Algorithm for Work in this Thesis

In this thesis, four layered depth images were constructed using four of the camera views and depth maps of the breakdancing scene from [5]. The rendering algorithm used for all LDIs from herein is given in Fig. 3.15.

In order to perform filtering of the LDI representation, we first review the characteristics of 2D spatial filtering in the following chapter.

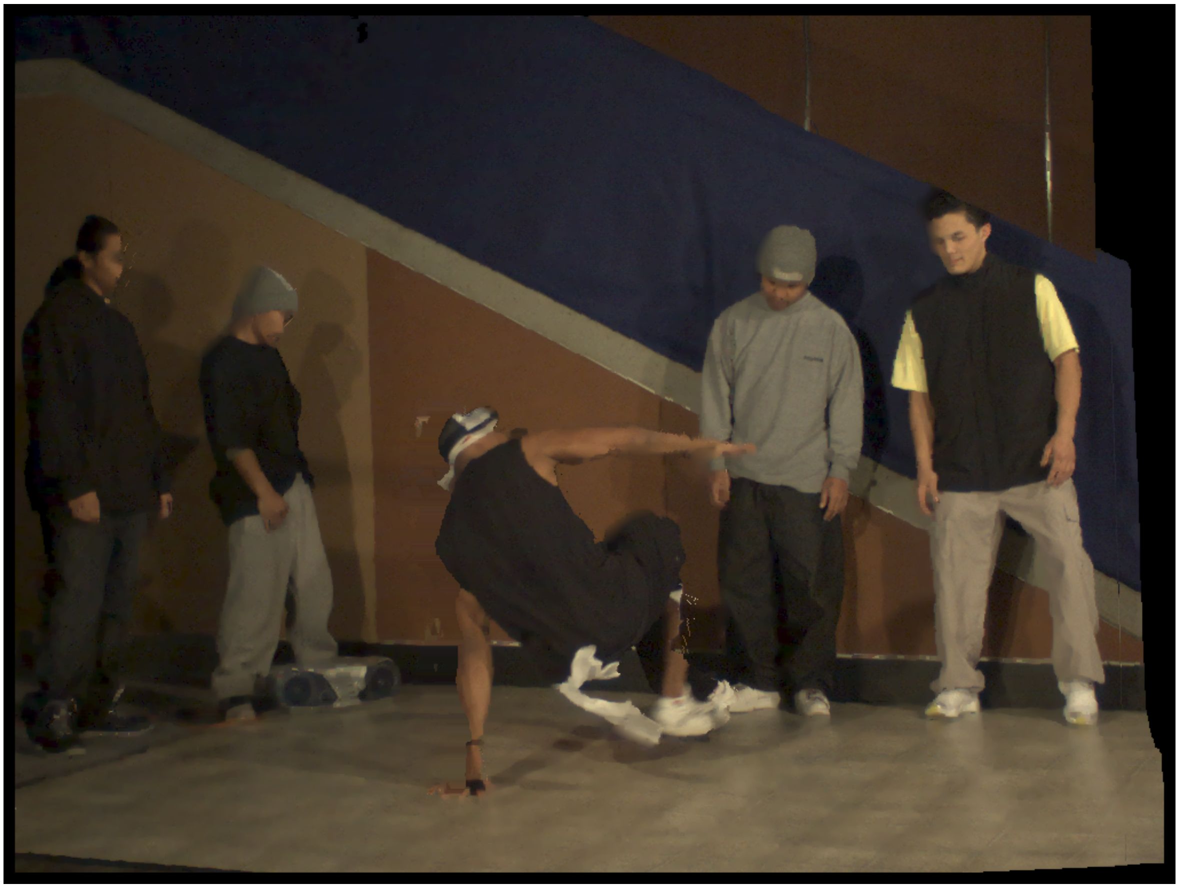


Figure 3.14: Large hole filling example using the method of [49]. The input image was that of Fig. 3.9(b) and the source region was set using the mask in Fig. 3.13(c).

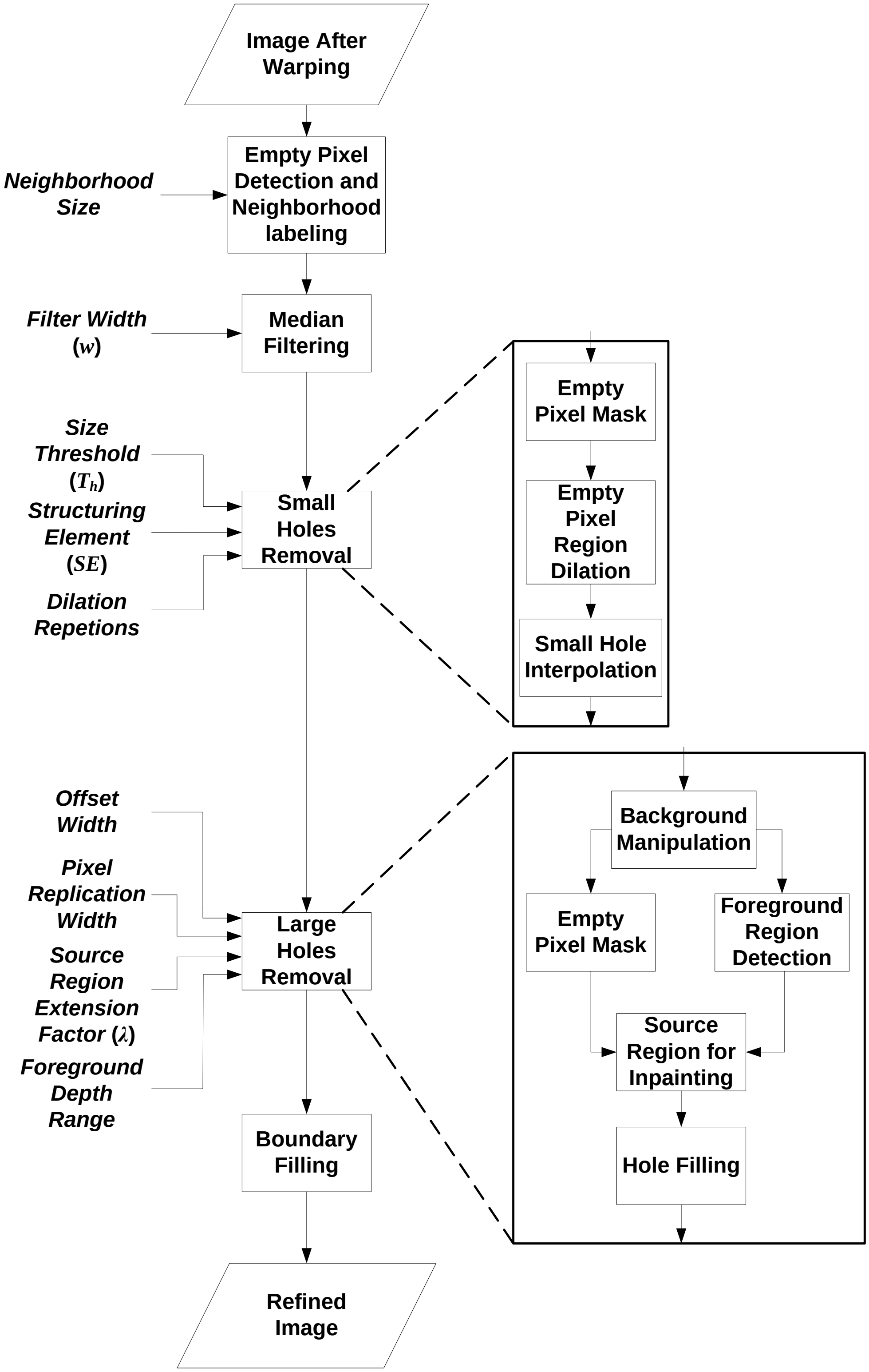


Figure 3.15: LDI rendering algorithm.

Chapter 4

2D Image Filtering

The majority of stereoscopic displays that exist today (in the year 2011) are 2D screens capable of presenting 2D images. The stereoscopic effect is obtained by directing the appropriate 2D image to each of the viewer's eyes using either a special type of lens placed above the display, or by wearing special glasses (please refer to Chapter 2 for more details). Since these images are typically stored/transmitted as 2D image pairs, the basic method to process these images is through use of 2D image processing techniques applied to the left and right images separately.

Images are processed for many reasons including analysis, and for enhancement to improve image re-production. When images are captured by a camera, they are subject to various forms of noise emanating from, but not limited to, the camera sensor, lens distortions, and low light conditions. Additionally, these images can appear out of focus if the camera focal length is not properly adjusted for the scene being captured. These types of problems can be corrected by using signal processing techniques achieved through filtering the images in either the frequency domain or by using linear spatial filtering in the spatial domain since they have a one-to-one correspondence.

When considering the choice of using spatial/frequency filtering for stereoscopic images, the method best suited to filtering layered depth images is spatial filtering since we can implement nonlinear filtering operations (which wouldn't be possible using a frequency domain filter). We therefore focus the remainder of the discussion in this chapter on 2D spatial filtering. Specifically, the following chapter discusses 2D spatial filtering and some of the common filters used for image smoothing and sharpening.

4.1 Basics of Linear Spatial Filtering

Filtering a 2D image $f(x, y)$ by a 2D spatial filter $w(x, y)$ of size $m \times n$ can be represented as in [16] by:

$$g(x, y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s, t) f(x + s, y + t), \tag{4.1}$$

where $g(x, y)$ is the filtered image and $a = (m - 1)/2$ and $b = (n - 1)/2$.

The process of spatial filtering is usually accomplished by zero-padding the border of the image with $m - 1$ rows, and $n - 1$ columns. Subsequently, the filter $w(x, y)$ traverses the image by visiting each pixel and computing the result $g(x, y)$ using (4.1). The collection of all individual results $g(x, y)$ are then combined to form the final result, which is the filtered image.

In order to achieve desired filter results, the filter coefficients in $w(x, y)$ need to be adjusted such that they correspond to the type of filter required. A description of image smoothing and sharpening (two common filtering operations), along with some of the common filters used to perform these operations follow.

4.1.1 Smoothing Filters

Image smoothing (also referred to as blurring) is a process by which the high frequency components of an image are de-emphasized. In a spatial filtering context, this process occurs by taking an average of pixels inside the filter window $w(x, y)$. Smoothing or low pass filtering (which is the suppression of high frequency components) is achieved in this manner since smooth regions of an image (i.e. where intensity changes are gradual or very minimal, like DC components) will generate a result which is very similar to the majority of the pixels in this region. In regions with greater variations of intensity change, these filters may produce results that are drastically different than the center point in the window itself, resulting in a very different pixel value (which is the suppression or blocking of high frequency regions of the image) and replaces them with less sharp intensity transitions.

Two examples of spatial smoothing filters that function in this manner are the *box filter* and the *weighted average filter*.

4.1.2 Box Filter

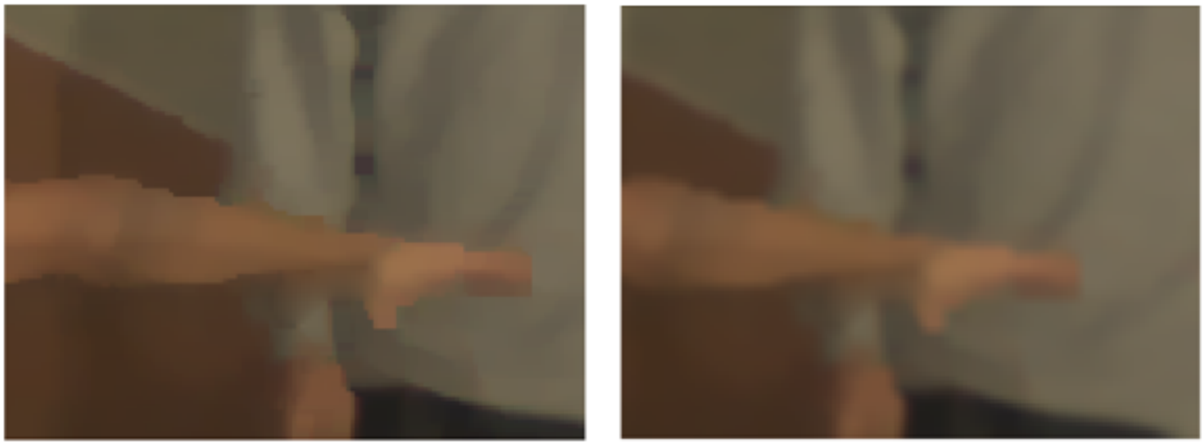
To smooth a 2D image, a common low pass filter (LPF) that is used is known as the box filter (or smoothing filter). The coefficients used in a 3×3 mask of this kind of LPF are,

$$\begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \times \frac{1}{9}, \tag{4.2}$$

and the result of applying this filter from [16] onto a 2D image is shown in Fig. 4.1(b).

The 2D box filter can achieve a very coarse blur for an image since the filter result is equally dependent on all pixels in the window. This type of filter is very primitive and significantly degrades the image as the filter dimensions increase. This occurs because all points neighbouring the center pixel are given an equal amount of weight in the filter result. Additionally, one can think that when a filter size is large, pixels that are distant from the center (i.e. near the perimeter of the filter window) are likely to be significantly different than the pixels nearest to the center. Thus, giving equal proportions of weight in the filter can significantly degrade quality in high frequency areas of an image.

When dealing with stereoscopic images, there is a 2D left eye image, and a 2D right eye image. By applying the 2D box filter to each view separately, the result produces



(a) Original Image from [5]

(b) Image (a) Filtered using 3 × 3 Box Filter

Figure 4.1: Effects of 2D Box Filter.

a blurring between foreground and background objects (see edges of the man's arm in Fig. 4.1(b). The problem with using this technique is that pixels in the background should not be filtered with the pixels in the foreground because the resulting image will show inconsistencies where pixels appear to have incorrect depth locations. The resulting image may lead to unnatural visualization. It is believed that spatial inconsistencies are a factor leading to visual discomfort [20].

To correct this problem, a wiser choice for a smoothing filter would need to be used where the effects of background pixels are minimized when filtering pixels in the foreground.

4.1.3 Weighted Average Filter

In order to decrease the amount of coarse blurring that occurs in high-frequency areas, the filter in (4.2) can be modified such that pixels closer to the center of the mask are given a greater relative weight in the filter result than pixels further away. Namely, (4.1) can be modified in this fashion to become a weighted average filter. The general form of a 2D weighted averaging filter of size $m \times n$ is described by [16] as,

$$g(x, y) = \frac{\sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t) f(x+s, y+t)}{\sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s,t)}, \tag{4.3}$$

where the denominator in (4.3) is the sum of the mask coefficients and is constant. This result can also be written as,

$$g_{WA}(x, y) = \frac{g(x, y)}{w(s,t)}, \tag{4.4}$$

where $g(x, y)$ is from the definition in (4.1) and $w(s,t)$ is taken as the sum of mask coefficients. A 2D weighted average filter mask where the center pixel has approximately 27 percent of the filter weight can be represented by,

$$\begin{array}{|c|c|c|} \hline 1 & 3 & 1 \\ \hline 3 & 6 & 3 \\ \hline 1 & 3 & 1 \\ \hline \end{array} \times \frac{1}{22}, \tag{4.5}$$

An example of an image filtered using this approach is shown in Fig. 4.2(d) where the filter mask used is shown in (4.5). A comparison of the filtering result of the 2D box filter and a 2D Weighted Average filter is shown in Fig. 4.2. From Fig. 4.2, it is obvious that the two filters achieve the same results in areas of zero intensity change. However, in regions of high frequency (such as the edges of the man's hat and face with the background), we notice that the box filter achieves more blur since the center pixel does hold more weight in the overall result. In terms of PSNR, the box filter (when compared to the original) was 37.2481 dB and the weighted average filter was 38.2992 dB. The resolution of the original image was 1024 × 768 pixels.

When considering the use of a smoothing filter for stereoscopic images, an important consideration when blurring the entire image would be to maintain consistent blurring in regions of similar depth. Generally, in 2D images, components that are in the foreground are typically in focus, while components of the background are out of focus. When using a box filter, blurring the boundaries of foreground objects increases the amount of influence by the background, which yields an image that is unnatural (also causing a mismatch when viewed stereoscopically, causing viewer discomfort). Instead of relying on purely 2D filtering techniques that have no knowledge of foreground and background, a better choice would be to use the weighted average filter since pixels on the boundaries will have higher relative weight than pixels in the background of an image. This will help to maintain the separation of foreground and background over the use a of box filter and reduce image inconsistencies where pixels appear to have incorrect depth locations (where the resulting stereo image may lead to unnatural visualization). However, the improvement in separately filtering the foreground and background is minimal, to say the least. Some knowledge of image depth would be required to influence the outcome of the filter, or a process of edge-preserving smoothing.

4.2 Bilateral Filter

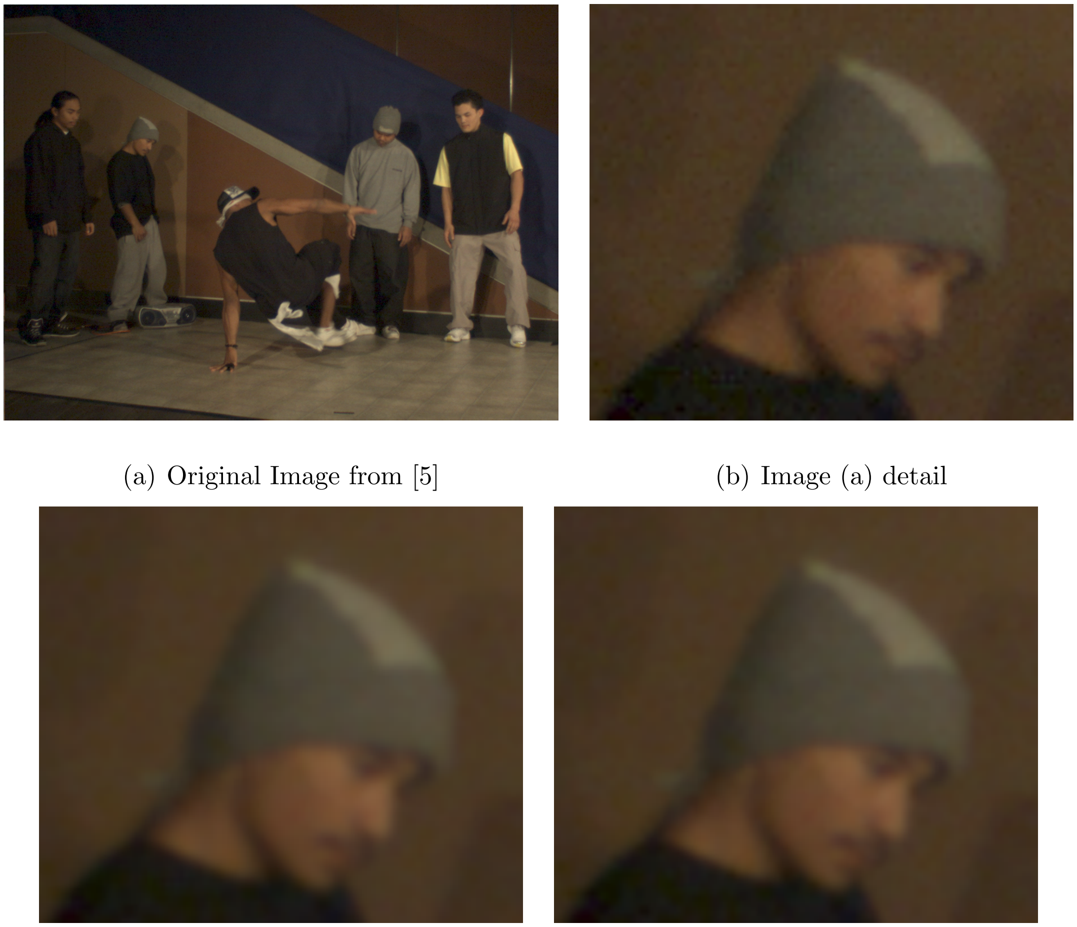
Edge-preserving smoothing can be accomplished by using a 2D bilateral filter as described in [50] by:

$$BF[I]_{\mathbf{p}} = \frac{1}{W_{\mathbf{p}}} \sum_{\mathbf{q} \in S} G_{\sigma_s}(\|\mathbf{p} - \mathbf{q}\|) G_{\sigma_r}(I_{\mathbf{p}} - I_{\mathbf{q}}) I_{\mathbf{q}}, \tag{4.6}$$

where $\mathbf{p}$, and $\mathbf{q}$ are pixel positions in the gray-scale image I, S is the set of possible positions in the image, and where the normalization factor $W_{\mathbf{p}}$ is defined as:

$$W_{\mathbf{p}} = \sum_{\mathbf{q} \in S} G_{\sigma_s}(\|\mathbf{p} - \mathbf{q}\|) G_{\sigma_r}(I_{\mathbf{p}} - I_{\mathbf{q}}). \tag{4.7}$$

In (4.6), the term $G_\sigma(x)$ represents the two-dimensional Gaussian kernel and G_{σ_s} is a spatial Gaussian that decreases the influence of distant pixels. Moreover, the choice of σ_s, which dictates the width of the Gaussian kernel, adjusts the relative size of image features to be blurred (i.e. the greater the width, the larger the feature that can be blurred). The bilateral filter can also preserve edges because the term G_{σ_r} is a range



(a) Original Image from [5]

(b) Image (a) detail

(c) Image (a) detail using 2D Box Filter

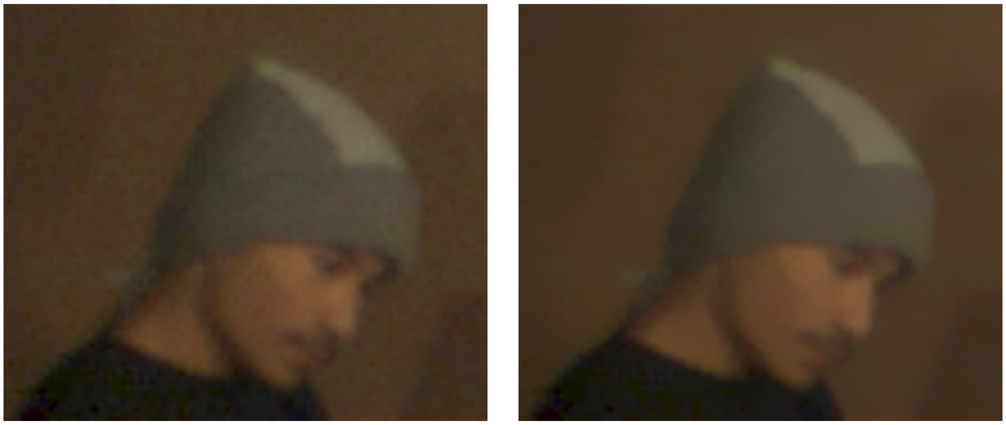
(d) Image (a) detail using 2D Weighted Average Filter from (4.5)

Figure 4.2: Comparison of 2D linear spatial filtering.

Gaussian that decreases the influence of pixels with different intensities from that of the centre pixel $I_{\mathbf{p}}$ [50].

An example of a 2D bilateral filter applied to an image is shown in Fig. 4.3(b). As shown in the image, there is a clear distinction of the foreground and background since the edges of the image have been preserved. A comparison of these 2D smoothing filters (box, weighted average, and bilateral) are shown in Table 4.1. Please note that each of the filters in Table 4.1 were implemented in brute force and without any optimization.

Although the 2D bilateral filter is a good choice for 2D images, it is not ideal for a multiview system. Specifically, when used in a multiview system, to achieve a smoothing



(a) Fig. 4.2(a) detail

(b) Image (a) detail after 2D Bilateral Filter of size 9 × 9, $\sigma_r = 4$ and $\sigma_s = 4$

Figure 4.3: Effects of 2D bilateral filtering.

Table 4.1: Comparison of 2D smoothing filters.

Smoothing Filter	**Window Size**	**PSNR (dB)**	**Processing Time (sec)**
Box size	3×3	37.2481	3.1
Weighted Average	3×3	38.2992	3.1
Bilateral	9×9	44.2577	76.1

of all possible views will require the bilateral filter to be applied separately to each possible image view (increasing the required processing time). If the multiview system is capable of rendering n views, then the 2D bilateral filter will need to be executed *2n* times since each view has a left and right eye image.

To alleviate the requirement of filtering each view separately, this thesis presents an adaptive 3D spatial filter that is designed to take advantage of the LDI representation to achieve the desired filtering result. The following chapter introduces this filter.

4.3 Sharpening Filters

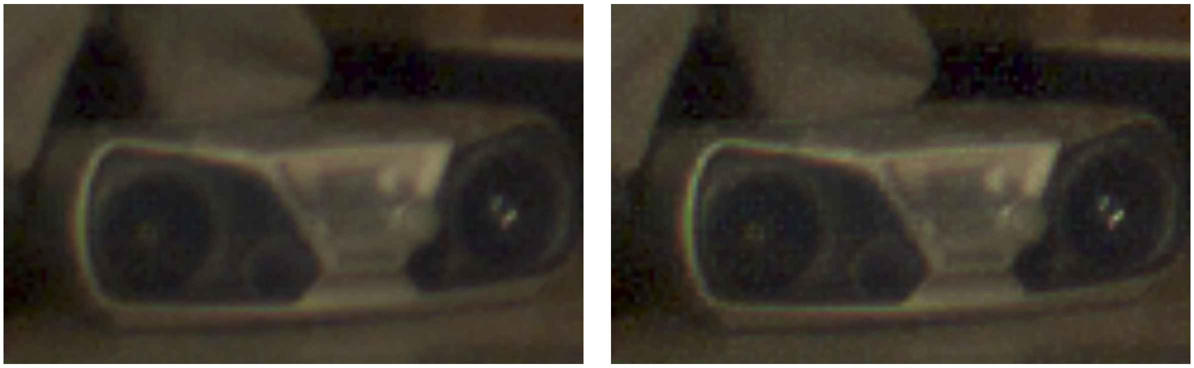
When images appear out of focus or blurry, they are over emphasizing the low frequency components of the image. Images captured by camera are usually blurry when the lens is not properly focused for the scene being captured. To correct the problem, the family of spatial sharpening filters can be used. Two of these common sharpening filters are described in the subsections below.

4.3.1 Unsharp Masking

In 2D, image sharpening can be achieved by subtracting the unsharp (i.e. smoothed) version of an image $b(x, y)$ from the original image $f(x, y)$. The sharpened image $s(x, y)$ can then be expressed as:

$$s(x, y) = f(x, y) + k[f(x, y) - b(x, y)], \tag{4.8}$$

where k is a weight that is set to 1 for unsharp masking [16]. The result of applying this filter onto a subsection of Fig. 4.2(a) is shown in Fig. 4.4, where the box filter was used to obtain $b(x, y)$. When comparing the images in Fig. 4.4, the edges and details in the stereo have been emphasized in Fig. 4.4(b). Specifically, the contours of the speakers inside the box are more obvious. In addition, since image sharpening enhances the high-frequency components of the image, the noise in the image has also been accentuated.



(a) Subsection of original image in Fig. 4.2(a)

(b) Subsection after unsharp masking

Figure 4.4: Effects of 2D unsharp masking.

4.3.2 Laplacian

Another sharpening filter that can be used to emphasize the edges in an image is the Laplacian filter. As described in [16], the Laplacian can perform sharpening since it implements a derivative operation. The derivative is used for digital sharpening because when the intensity of an image is constant, the derivative is zero, and when the intensity is changing, the derivative is non-zero - which is the basis of image sharpening. This can be done as in [16] by,

$$g(x, y) = f(x, y) + \nabla^2 f(x, y), \tag{4.9}$$

where $g(x, y)$ is the sharpened image and $f(x, y)$ is the original image. The term $\nabla^2 f(x, y)$ can be implemented using a mask (from [16]) such as,

$$\begin{array}{|c|c|c|} \hline 0 & -1 & 0 \\ \hline -1 & 4 & -1 \\ \hline 0 & -1 & 0 \\ \hline \end{array}\,, \tag{4.10}$$

which must then traverse the image as described in Section 4.1.

Chapter 5

3D Spatial Filtering for Multiview Images

An extension of the 2D linear spatial filter into 3D can be achieved by adding a third dimension, the z-axis, to (4.1). Specifically, the 3D FIR spatial filter $w(x,\ y,\ z)$ of size $m \times n \times p$ when applied to a 3D image $f(x,\ y,\ z)$, which in our case is the LDI, can be represented now as:

$$g(x, y, z) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} \sum_{u=-c}^{c} w(s, t, u) f(x + s, y + t, z + u), \tag{5.1}$$

where $g(x,\ y,\ z)$ is the filtered image (or LDI), $a = (m - 1)/2$, $b = (n - 1)/2$ and $c = (p - 1)/2$. A filter $w(x, y, z)$ with size $m \times n \times p$ and associated coefficients $w_{x,y,z}$ is shown in Fig. 5.1.

5.1 Explanation (Mechanics)

To perform spatial filtering on an LDI, one must follow the same mechanics as those required for a 2D image as stated in [16] with the addition of zero padding and traversing the filter in the z-direction. When performing 3D spatial filtering on an LDI, p number of layers are engaged in each calculation. An example of these mechanics are shown in

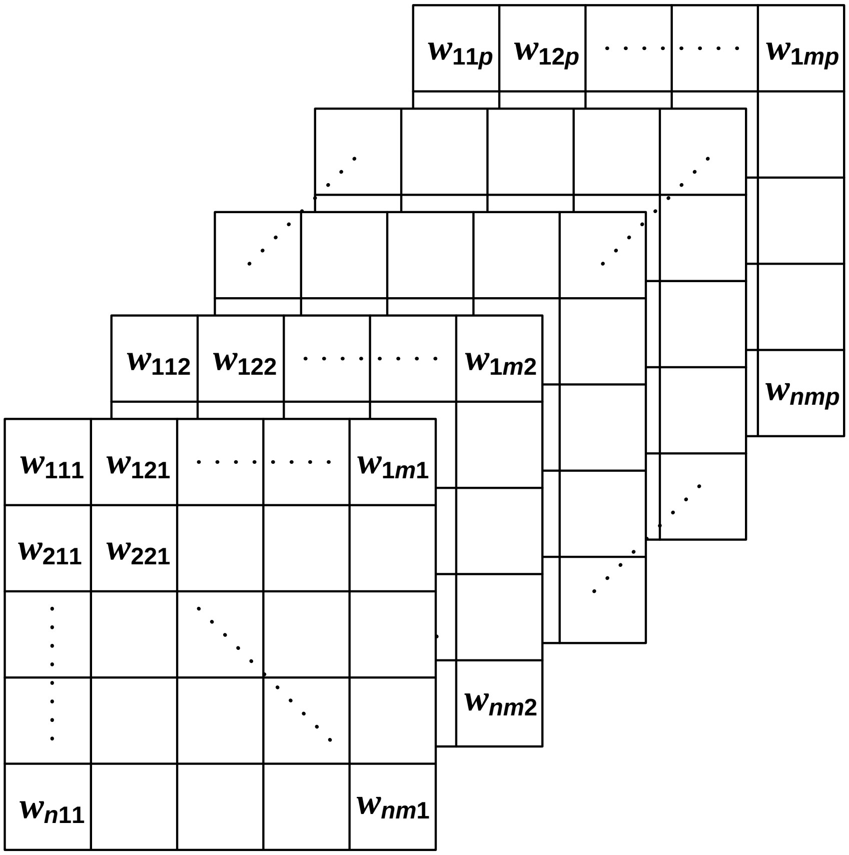


Figure 5.1: A general 3D spatial filter $w(x, y, z)$ of size $m \times n \times p$.

Fig. 5.2.

5.2 3D Smoothing Filter

A basic filter that can be used to smooth an LDI using (5.1) would be the extension of the 2D box filter in (4.2) to 3D. Specifically, the 3D version of this mask would be,

$$w(:,:,-1) = \frac{1}{27} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad w(:,:,0) = \frac{1}{27} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad w(:,:,+1) = \frac{1}{27} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \quad (5.2)$$

The extension of the 2D box filter to 3D in (5.2) is not ideal for smoothing the LDI

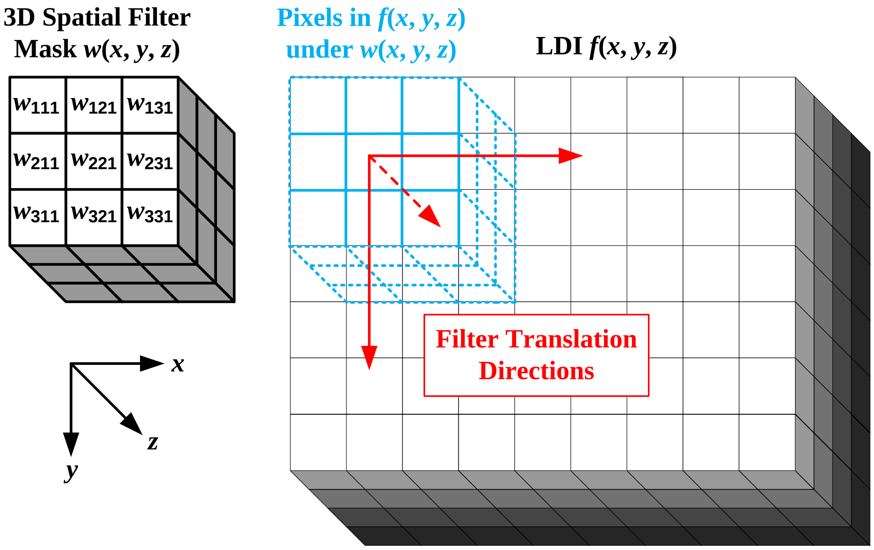

3D Spatial Filter Mask w(x, y, z)
w111
w121
w131
w211
w221
w231
w311
w321
w331
Pixels in f(x, y, z) under w(x, y, z)
LDI f(x, y, z)
Filter Translation Directions
x
y
z

Figure 5.2: Mechanics of 3D spatial filtering.

since the resolution of the LDI in the z-direction is much less than those of x and y. This will lead to uneven/unnatural blurring across depths, since successive depth pixels in the LDI with the same $(x,\ y)$ coordinate may have a large difference in their true z coordinates. Thus, a better smoothing filter to use in this case would be a weighted average filter that gives more weight to pixels residing in the center plane. The general form of a 2D weighted averaging filter of size $m \times n$ was shown in (4.3). The extension of this filter to 3D would lead to,

$$g(x,y,z) = \frac{\sum_{s=-a}^{a} \sum_{t=-b}^{b} \sum_{u=-c}^{c} w(s,t,u) f(x+s,y+t,z+u)}{\sum_{s=-a}^{a} \sum_{t=-b}^{b} \sum_{u=-c}^{c} w(s,t,u)}. \tag{5.3}$$

The denominator in (5.3) is the sum of the mask coefficients and is constant. This straightforward extension would be ideal for smoothing a 3D volume where each image slice has equal resolution. However, in the case of the LDI, the resolution in the residual layers becomes more sparse for deeper layers. Consequently, filtering over pixels that are empty can lead to a result that is invalid since not every pixel in the neighborhood will be defined.

5.2.1 Adaptive Qualities of the Filter

In order to be able to use the extension of (4.3) in 3D for LDIs, the size of the mask would need to become adaptive to accommodate the varying distribution of layered depth pixels. Specifically, since not all layered depth pixels have depth pixels in each plane (see Fig. 5.3), the filter mask would need to change size based on the number of non-empty pixels present for each depth plane. Moreover, the number of depth pixels is usually dependent on the complexity of the scene where the higher the complexity, the greater the amount of depth pixels. We can then define the number of non-empty pixels in an arbitrary depth plane $z = i$ by $\beta_{z=i}$. An example of this adaptive nature is shown in Fig. 5.4.

For one to better understand this concept of $\beta_{z=i}$, an example is shown using a 3D weighted average filter (where the center plane has a 60% weight) as shown in Fig. 5.5. When considering the mask in Fig. 5.5, the sum of the valid mask coefficients (that have defined pixels in the LDI structure) and the size of the mask is dependant on $\beta_{z=i}$. Therefore, the denominator of (5.3) would be equal to,

$$\sum_{s=-a}^{a}\sum_{t=-b}^{b}\sum_{u=-c}^{c} w(s,t,u) = \left(\frac{2}{\beta_{z=i}}\right)\beta_{z=i} + \left(\frac{6}{\beta_{z=i+1}}\right)\beta_{z=i+1} + \left(\frac{2}{\beta_{z=i+2}}\right)\beta_{z=i+2} = 10, \quad (5.4)$$

which is the sum of all mask coefficients that have existing pixels in the LDI. We can use the following to achieve a 3D filtered LDI using a weighted average filter from Fig. 5.5,

$$g_{WA}(x,y,z) = \frac{g(x,y,z)}{\sum_{s=-a}^{a}\sum_{t=-b}^{b}\sum_{u=-c}^{c} w(s,t,u)} = \frac{g(x,y,z)}{10}, \quad (5.5)$$

where $g(x,y,z)$ is from the definition in (5.1). Note that when the pixels of a specific layer under the filter are all empty (meaning that the layer does not contain any information in this part of the image or $\beta_{z=i} = 0$), the result is dropped from the calculation of the filtering result. Instead, an average of the remaining layers is taken to carry out

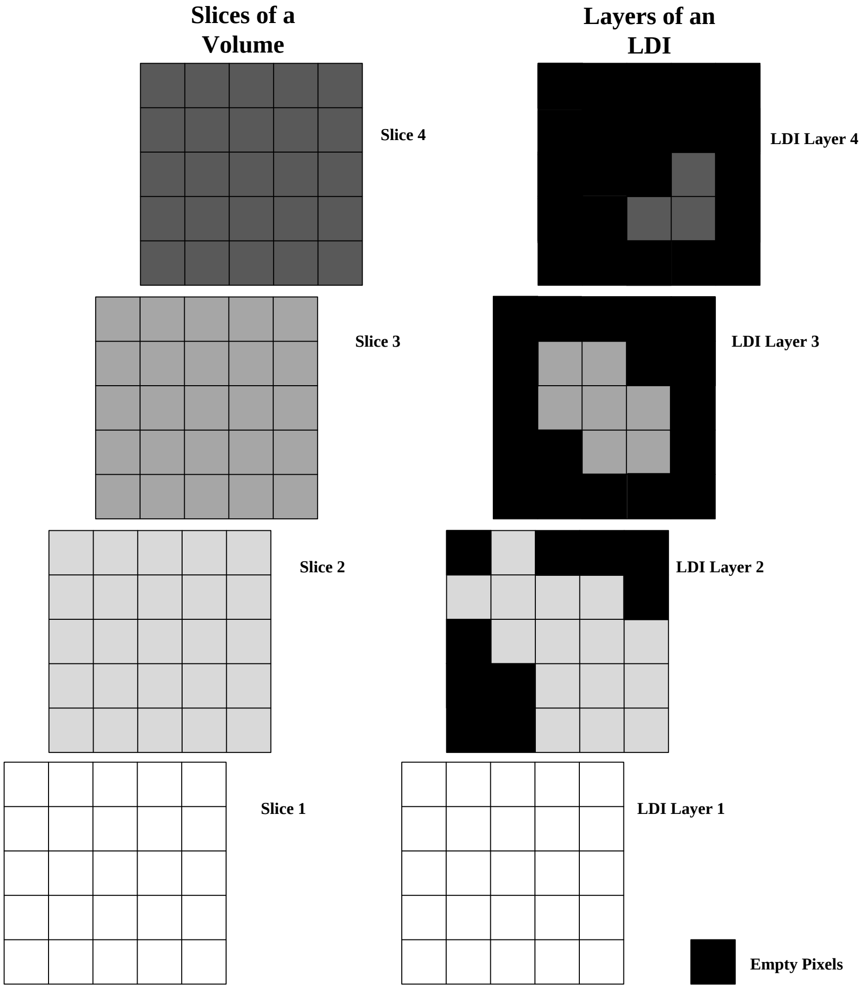


Figure 5.3: Sparse nature of an LDI compared to a ideal volume.

the filtering calculation. Subsequently, the denominator in (5.3) would need to change based on the new calculation of $w(s,\ t,\ u)$ with the appropriate planes dropped. If this adaptation is not taken into account, then the straightforward extension of a 3D spatial filter for the LDI will be inaccurate since depth pixels that may not exist in space will be allocated a weight in the calculation (which may yield a resulting image with invalid colours).

After the 3D filtering is completed in the "layered depth domain", the left and right views can then be rendered using the techniques described in Chapter 3.

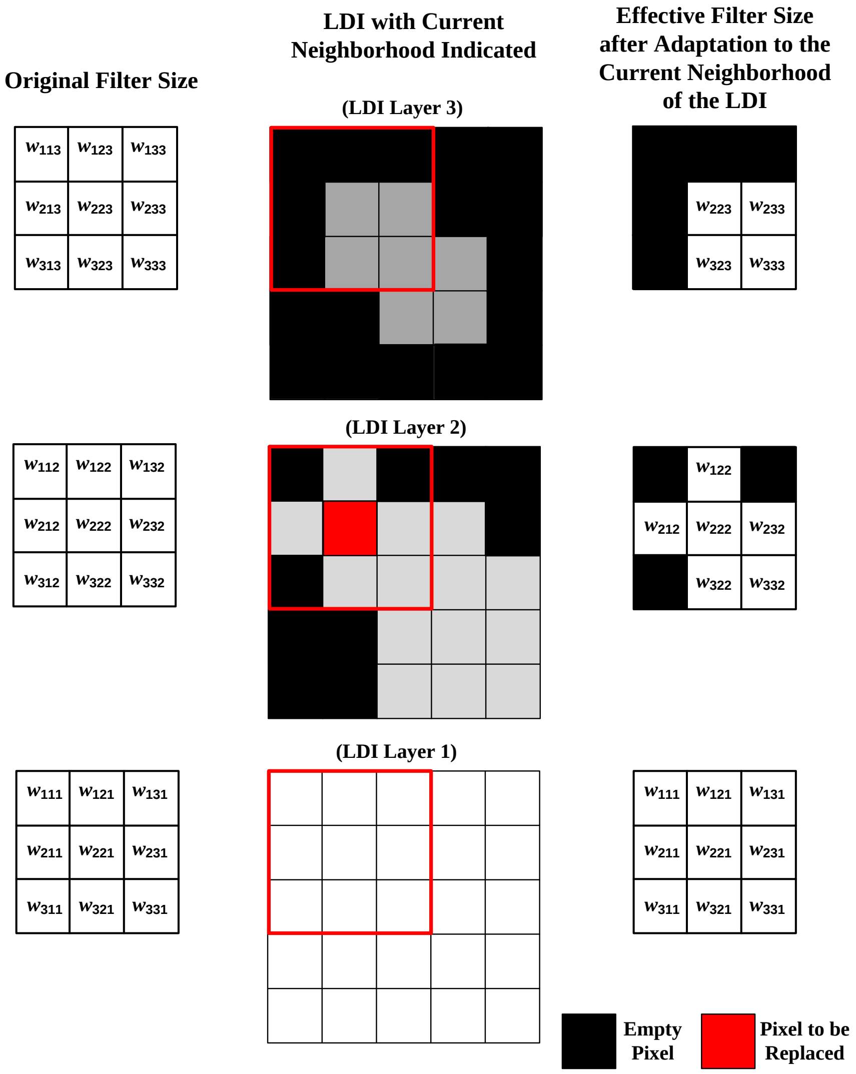
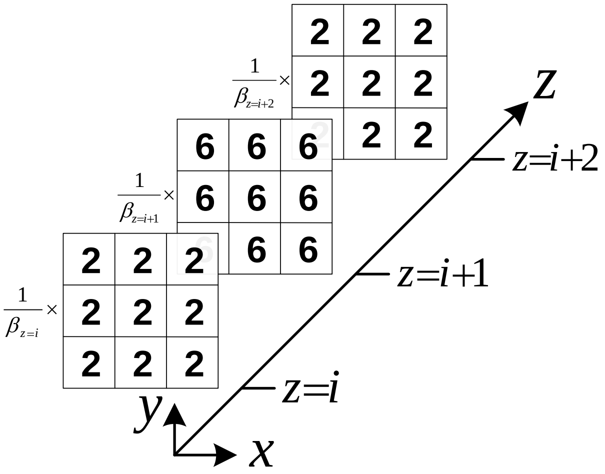


Figure 5.4: Example of the adaptive nature of the proposed filter. The example uses a filter of size 3 × 3 × 3.

5.3 Human Perception Experiments

The following section describes the experiment used to compare the differences in human perception between rendering smooth multiview images using either the 3D filter proposed or 2D bilateral filters, as in [51].



$\frac{1}{\beta_{z=i+2}}\times$
2 2 2
2 2 2
2 2 2
$\frac{1}{\beta_{z=i+1}}\times$
6 6 6
6 6 6
6 6 6
$\frac{1}{\beta_{z=i}}\times$
2 2 2
2 2 2
2 2 2
Z
z=i+2
z=i+1
z=i
y
x

Figure 5.5: Three-dimensional weighted average mask.

5.3.1 Subjects

A total of 15 subjects participated in the preliminary stages of this study. Their mean age was 24 years old. All of the subjects participating in the study had normal visual acuity and normal binocular depth perception. None of the participants were feeling ill, dizzy, or sick prior to the experiment.

The participants were unaware of the purpose of the experiment. They were told that they were participating in a study to assess the naturalness and level of viewing discomfort experienced when viewing stereoscopic images. They were told that the findings would help to better understand viewing discomfort and further knowledge into creating more natural images.

5.3.2 Test Images

Stereoscopic images from LDI data were rendered using both the filtering techniques of the 2D bilateral filter and the proposed 3D filter. A total of fifteen stereoscopic image pairs were shown to each of the subjects. Of the 15 pairs, five of them were filtered using a 3D filter, five were filtered with the 2D bilateral filter using a $\sigma_r = 0.05$, and the other five were filtered with the 2D bilateral filter using a $\sigma_r = 0.2$. The full list of filter specifications used in the experiment are shown in Table 5.1. In Table 5.1, the "Level" represents the relative weight of the center pixel for the 2D filters. For the 3D filter, the parameter CW represents the relative weight of the central plane with respect to the other planes. These parameters were set according to the relative "Level" used for the 2D filters (for comparative purposes).

Specifically, for every given "Level", the average percent difference between the weight of each coefficient in the 3D mask and the coefficient of the central pixel was set similarly to the percent difference of the coefficients in the counterpart 2D filter. It is worth noting that the 3D mask in "Level 1" indicates a 3D box filter, whereas the 3D mask in "Level 5" indicates a weighted average filter where almost the entire weight of the calculation depends on the center plane. In addition, all filters used in the experiment had a window size of 3×3 for the 2D filters tested, and $3 \times 3 \times 3$ for the 3D filters tested. An example of two stereo image pairs used in the study are shown in Fig. 5.6 and Fig. 5.7.

The 15 stereo pairs were presented, using a *single-blind* method, to the users in a random order (using a random number generator to determine the order). These orders were obtained prior to testing and were only known to the investigators. The 15 stereo pairs were arranged into three distinct sequences (sequence A, B, C) so as to randomize the order of viewing so that users could not guess any pattern in the images they were shown. Five of the subjects saw sequence A, five others saw sequence B, and five others saw sequence C.

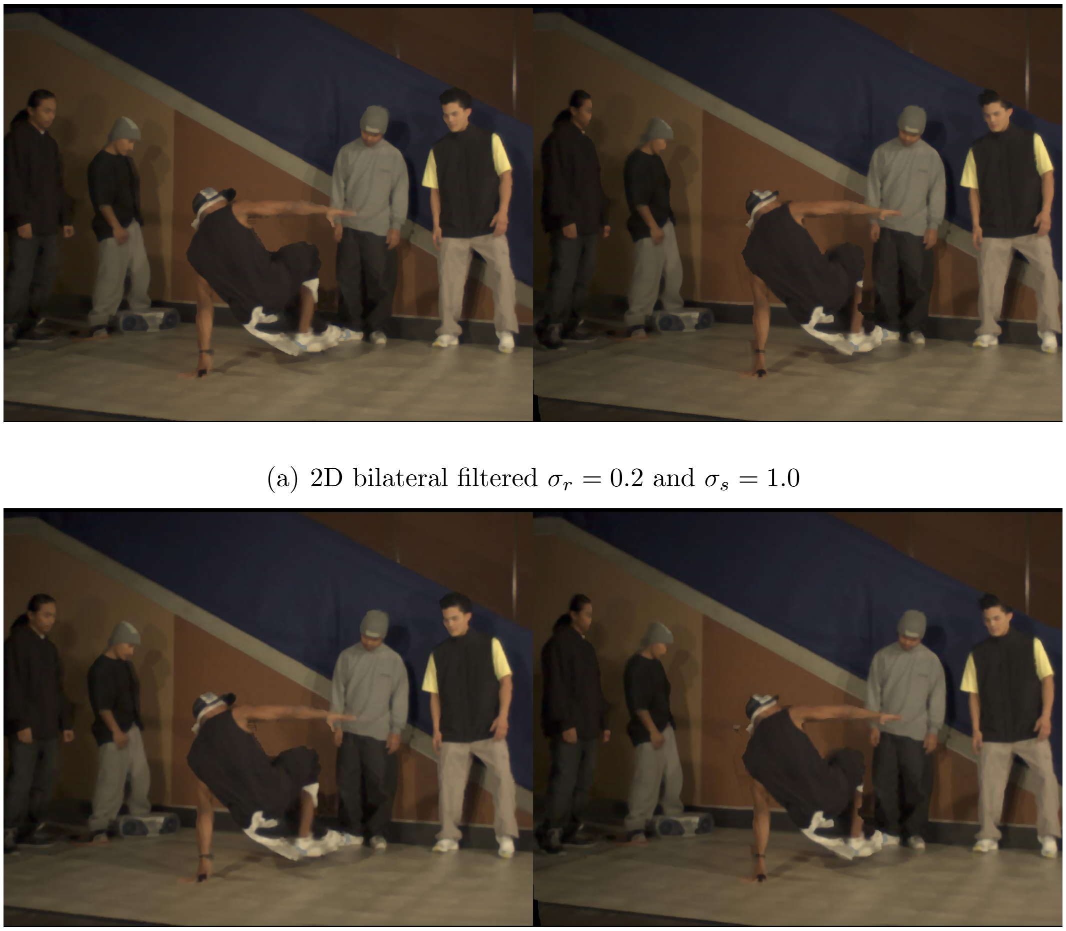
Table 5.1: Filter specifications used in the experiment (where CW represents the weight of the central plane with respect to the other planes).

Filter	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**
2D Bilateral ($\sigma_r = 0.2$)	$\sigma_s = 10$	$\sigma_s = 1.5$	$\sigma_s = 1$	$\sigma_s = 0.8$	$\sigma_s = 0.6$
2D Bilateral ($\sigma_r = 0.05$)	$\sigma_s = 10$	$\sigma_s = 1.5$	$\sigma_s = 1$	$\sigma_s = 0.8$	$\sigma_s = 0.6$
3D Proposed	$CW = 0.33$	$CW = 0.5$	$CW = 0.6$	$CW = 0.71$	$CW = 0.8$

The test images used in this experiment were generated using four camera views of the breakdancing scene from [5]. The original camera view (shown in detail in Fig. 5.8) used as the reference view for the LDI was generated using camera view 4.

5.3.3 Experiment Design

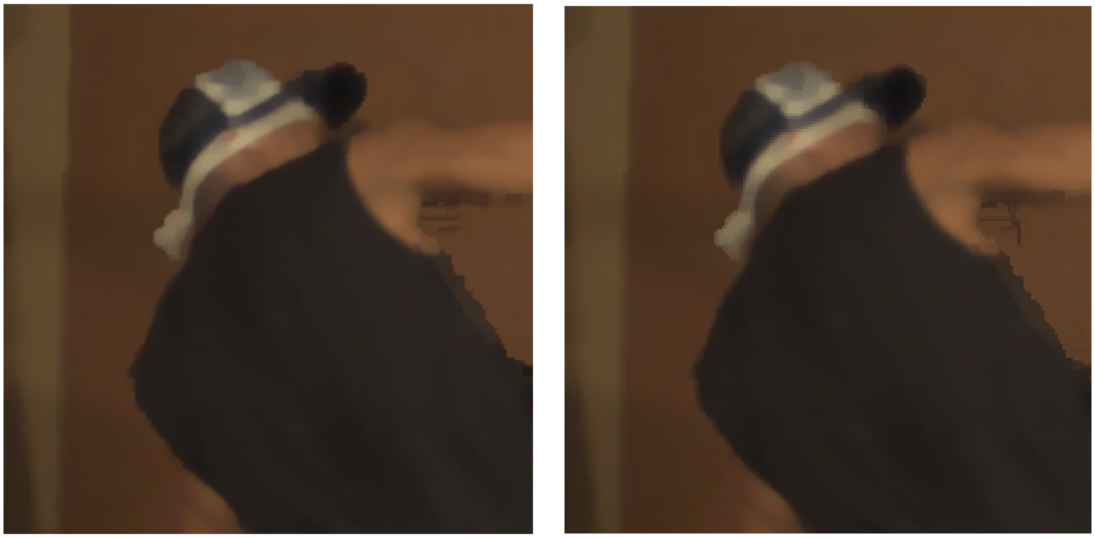
The subjects, investigated one at a time, sat in front of a 3D monitor at a viewing distance of 2.5 feet from the screen and wore special LCD shutter glasses. They were shown 15 images and asked to fill out a short questionnaire, based on the *Double-Stimulus*



(a) 2D bilateral filtered $\sigma_r = 0.2$ and $\sigma_s = 1.0$

(b) 3D filtered $CW = 0.6$

Figure 5.6: Example of filtered stereo pairs used in the experiment. These images will become available online at www.ryerson.ca/digitalmedialab



(a) Detailed result from Fig. 5.6(a)

(b) Detailed result from Fig. 5.6(b)

Figure 5.7: Example of detailed filter results (right views).

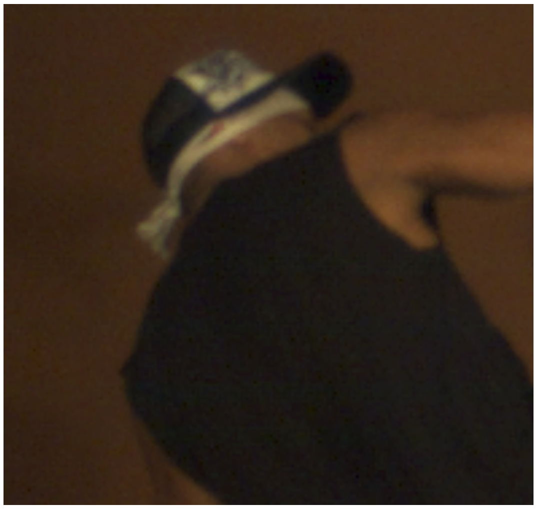
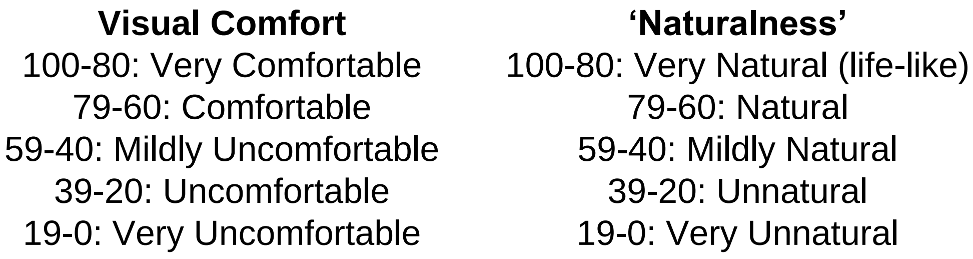


Figure 5.8: Detail of original image used for LDI reference view.

Continuous-Quality Scale protocol from the International Telecommunication Union recommendations (ITU-R BT.500-11), throughout the duration of the experiment. The experiment time line was:

1. Subjects were shown a reference 3D image that had no filtering applied to it (for six seconds).



Visual Comfort	'Naturalness'
100-80: Very Comfortable	100-80: Very Natural (life-like)
79-60: Comfortable	79-60: Natural
59-40: Mildly Uncomfortable	59-40: Mildly Natural
39-20: Uncomfortable	39-20: Unnatural
19-0: Very Uncomfortable	19-0: Very Unnatural

Figure 5.9: Rating scales for visual comfort and 'naturalness'.

2. Subjects were then shown a grey screen (for six seconds).

3. Subjects were then shown an image announced as "image one" (for six seconds).

4. Subjects were then shown a grey screen (for six seconds).

5. Subjects then had 10 seconds to fill out their scores for their perceived 'naturalness' of the images and the level of visual comfort by assigning a mark out of 100 using the respective scales shown in Fig. 5.9.

6. Steps 3 to 5 were then repeated in a similar fashion for the remainder of the images in the test sequence.

5.3.4 Equipment

The images were shown on a 23" *Alienware* 3D LCD display. The LCD shutter glasses used in the experiment were the pair included as part of *Nvidia's 3D Vision Kit.*

5.3.5 Experimental Conditions

The images were shown in a room that had control of the ambient light and ambient noise.

The filter described in (5.3) can also be easily used to perform sharpening of a stereoscopic image. This extension is discussed in the following subsection.

5.4 3D Sharpening Filter

To apply the unsharp masking technique to a stereo image (using the LDI representation), the 3D weighted average filter of (5.3) can be used to smooth the LDI to obtain a smoothed version of each layer $b_{z=i}(x, y, i)$. Following this operation, the sharpened result of each layer in the image can be computed using,

$$s_{z=i}(x, y, i) = k(f_{z=i}(x, y, i) - b_{z=i}(x, y, i)) + f_{z=i}(x, y, i), \tag{5.6}$$

where $s_{z=i}(x, y, i)$ represents the sharpened version of layer $f_{z=i}(x, y, i)$ as in [52].

Chapter 6

Results and Evaluation

The following chapter discusses the results and evaluation of the proposed filtering method from Chapter 5 using acquired camera images. The original multiview images used to present the results in this chapter were obtained from [5]. This dataset was chosen since it encompassed two sequences with many frames in addition to the specification of extrinsic/intrinsic camera parameters and depth maps for each image.

It is worth mentioning that each of the images shown in this chapter were obtained by extracting one view from a multiview stereo image. The full stereoscopic images are available at www.ryerson.ca/digitalmedialab.

6.1 Smoothing Results

The images shown in Fig. 6.1 through Fig. 6.4 show a comparison of the proposed 3D smoothing filter with 2D image filters. Each of the images were generated using a four layer LDI. In Fig. 6.1 through Fig. 6.4, the 2D box filter had a 3×3 window size, the 2D bilateral filter used a 3×3 window size and a $\sigma_r = 4$ and $\sigma_s = 4$, while the 3D smoothing used the $3 \times 3 \times 3$ proposed filtering method for LDIs.

In Fig. 6.1, the 3D filtered image shows blurring of the original image, but a distinct separation between the foreground and background is maintained (see edge of the person's face and hand). In contrast, this separation is not possible if a 2D box filter were used for smoothing as shown in Fig. 6.2 (edge of person's face and hand are blurred with the background).

The 2D bilateral filter which solves this problem of separately filtering different components of the image is shown in Fig. 6.3 (which shows blurring of foreground and background while maintaining sharp edges). When comparing the images in 6.3, the differences in smoothing are very subtle where the main discernible difference is in the distinction of edges (where the bilateral filter is more pronounced). Lastly, the small artifacts in the images discussed were caused from inaccuracies in the depth maps and ghost contours. A second example image, with the same comparisons as in Fig. 6.1 through Fig. 6.3, is shown in Fig. 6.4.

6.2 Sharpening Results

The images shown in Fig. 6.5 and Fig. 6.6 show a comparison of the 3D sharpening method proposed with 2D unsharp masking. In both Fig. 6.5 and Fig. 6.6, the 2D unsharp masking used a 3×3 box filter to obtain the unsharp mask, while the proposed 3D sharpening method for LDIs used a $3 \times 3 \times 3$ filter (using the proposed smoothing method for LDIs) to obtain the unsharp mask.

In Fig. 6.5(b), the 3D filtered image shows sharpening of the original image shown in Fig. 6.5(a). Specifically, the high frequency details in the image have been amplified. For example, the edges between the person's fingers and other details (such as noise) are also more visible. This is similar and very hard to distinguish from the result obtained using the 2D unsharp masking technique as shown in Fig. 6.5(c). A second example image, with a similar comparison as in Fig. 6.5, is shown in Fig. 6.6.

6.3 Evaluation of 3D Smoothing (Comparison to the 2D Bilateral Filter)

6.3.1 Visual and Human Perception Results

After conducting the experiment described in Section 5.3, the data collected from each of the participants was compiled and matched to its respective filter. For each filter, the mean ratings for visual comfort and naturalness were then tabulated along with the standard deviations for the sample. A comparison of these ratings for the bilateral filters and 3D filters, along with their associated 95% confidence intervals, are shown in Fig. 6.7 and Fig. 6.8 for the visual comfort and naturalness, respectively. The 95% interval was calculated as was proposed in recommendation ITU-R BT.500-11.

Evaluating stereoscopic images subjectively is currently a common practise since measures such as *MSE*, or *SNR* (that are seldom used for 2D images) don't accurately reflect the quality of stereoscopic images. Furthermore, stereoscopic images are fused by the brain to give a sensation of depth, and high SNR does not necessarily mean the brain will have an easier time to perceive depth or interpret natural looking stereoscopic images. Therefore, it is more important to evaluate the presented work subjectively because it will reflect to a greater extent, the actual human perception of these stereoscopic images and the filtering methods.

The results in Fig. 6.7 and 6.8 show that the proposed 3D filter achieves comparable results for comfort and naturalness to the various levels of the bilateral filter. Specifically, for visual comfort, the proposed 3D filter compared to the 2D bilateral filter had an average percent difference of 7.15% for the sample. For naturalness, the average percent difference of the proposed 3D filter compared to the 2D bilateral filter was 7.31%. Overall, the worst difference for an individual mean was 14.6%, while the best case (i.e. the smallest difference between the mean of the 3D filter and the mean of the highest rated

2D bilateral filter) was 1.69%.

More over, the displayed confidence intervals show significant overlap between the proposed 3D filter and the 2D bilateral filters. When taking into account the small sample size, there is greater margin for error of where the “true” mean of the samples should lie. Therefore, as stated in ITU-R BT.500-11, this interval can predict with a probability of 95% that the absolute value of the difference between the true and actual mean score will be smaller than the 95% confidence interval. This indicates that the true perceived quality of the proposed 3D filter and the 2D bilateral filters by the entire population will likely fall into this interval.

A possible explanation for the worst case percent difference of 14.6%, where the proposed 3D filter did not perform as closely as the bilateral filters would be due to the relative weight of the center pixel/plane being set to ‘1’. In this case, the 3D filter is acting as a 3D box filter, which (as mentioned earlier) is not ideal for LDIs since the resolution in the z direction is not as high as the x and y. Therefore, giving each plane equal weight in the filtering result reduced the visual comfort perceived by the participants.

In terms of human perception, the bilateral filter with $\sigma_r = 0.05$ generally had the highest ratings for naturalness and visual comfort. This can be attributed because the difference between the original image and the filtered image was very small since the range of 5% preserved almost every edge in the image, while the 3D filter only preserved edges in and around depth regions.

The results also showed that, in terms of visual comfort, the average person tested preferred “Level 2” with respect to the specifications given in Table 5.1 for the 3D filter. This means that the optimal parameters for a 3D filter (for the group studied) would be similar to those in Fig. 5.5 but with the coefficients in the center plane equal to 4.

6.3.2 Computational/Cost

The time taken to perform filtering using a computer with an *Intel* i7 920 (2.67 GHz) processor for the 2D bilateral filter was 48.17 seconds (for one out of two views for a single stereo pair). The time taken to perform 3D filtering, capable of rendering any desired view thereafter, using the proposed 3D filter in the "layered depth domain" was 28.36 seconds. Both methods were implemented in brute force and the computation time was measured without considering the time required to estimate depth. These results are shown in Table 6.1.

The reason the proposed 3D spatial filter requires much less computation time than the 2D filter is because it only needs to be applied once while the data is in the LDI representation. Subsequently, any stereoscopic pair of views for multiview applications can be rendered encompassing the desired (filtered) characteristics. The overall algorithm used to filter an LDI with the 3D proposed filter is shown in Fig. 6.9(a). As shown, the filter is only applied once for n views. Specifically, for an LDI with resolution in the x and y directions of R, and a number of depth layers d, the proposed 3D filter of size $w \times w \times w$ would give a worst case complexity of $O(w^3 \cdot R \cdot d)$. However, since the nature of the LDI is sparse in residual layers (i.e. behind the first layer), we can expect the worst case computation time to be less than the one mentioned above.

On the other hand, the 2D spatial filter requires more computation time because it must be applied to each rendered view separately. Consequently, increasing the number

Table 6.1: Processing time comparison.

Filtering Method	**Processing time required for 1 image**	**Processing time required for 4 stereo pairs**
2D Bilateral	48.17 sec	385.36 sec
3D Proposed	28.36 sec	28.36 sec

of stereo pairs for multiview applications will require more processing time if the views are to be processed off-line. The overall algorithm used to filter an LDI with the conventional 2D spatial filter is shown in Fig. 6.9(b). As shown, the filter must be applied n times in order for all views to be filtered. Moreover, if we define the number of views by n, where n is two times the number of stereo pairs to be rendered (since each stereo pair has a left and right image), and if we let the resolution of each 2D image be R, we can then define the worst case complexity for applying the 2D bilateral filter to multiview stereo images as $O(n \cdot R \cdot w^2)$. This complexity is taken where w^2 is the neighborhood search size for the bilateral filter.

6.3.3 Application to Various Types of Images

When comparing the results of the proposed 3D filter with the bilateral filter, it is very hard to distinguish a difference when viewing acquired images. Subsequently, a viewer will have a hard time telling a difference when shown one of these images on a small stereoscopic screen (such as on a handheld device) or a computer monitor. Likewise, consumers typically are not always experts in imaging and cannot discern the differences in quality, which make the proposed filtering method applicable to small screens or handheld devices and applications which are not heavily dependant on quality. Also, for certain applications where processing speed is critical, the results of the experimental study show that using the proposed 3D filter will save time without compromising enough quality that the average person can detect. Some examples of real-time applications that could benefit from this work are telepresence on a cell phone or streaming multiview stereoscopic content onto a handheld device.

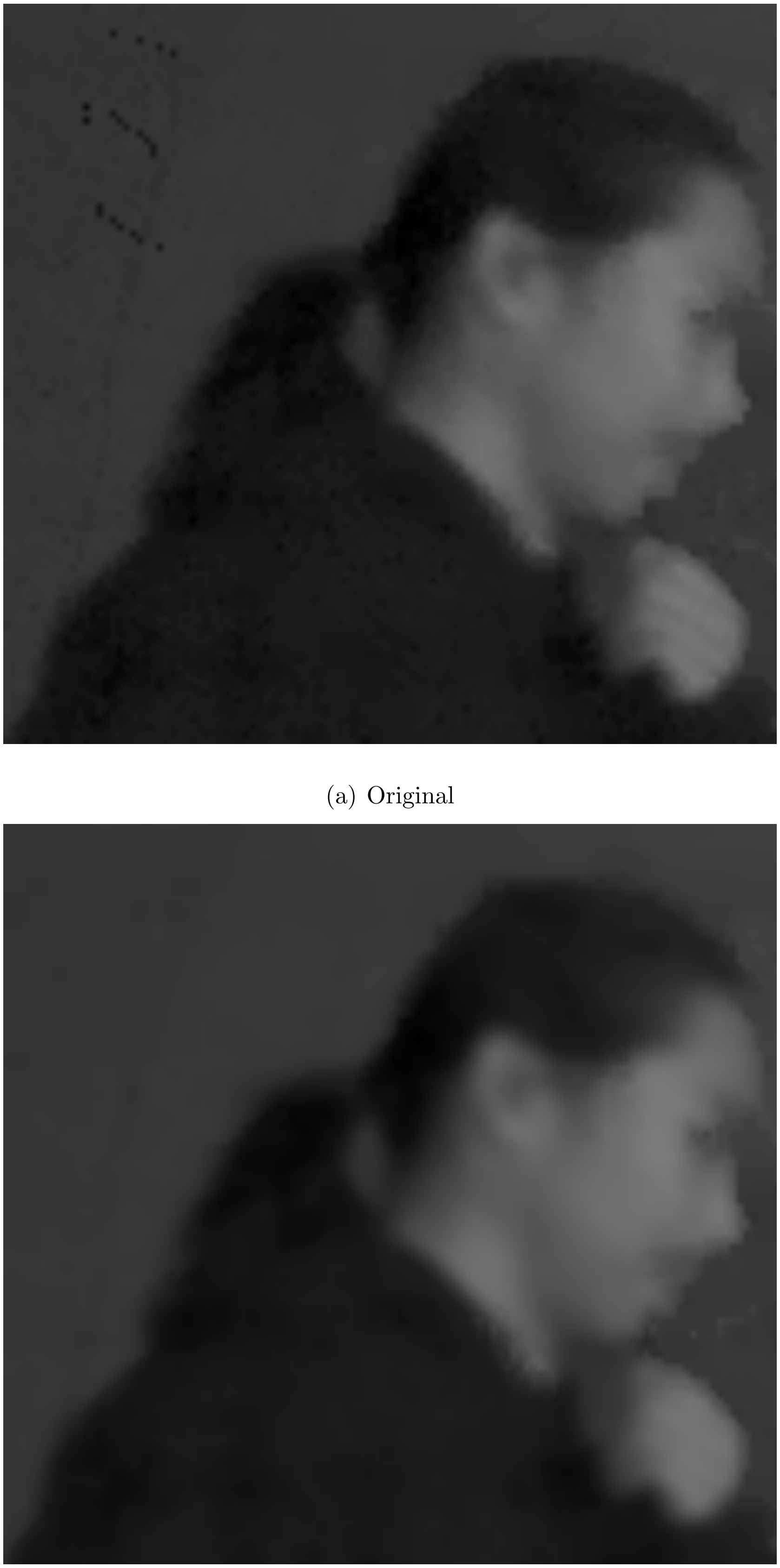
Alternately, in applications where the processing speed is neither critical nor required to be real time, but instead, more importantly required to have higher visual quality and accuracy (such as digital cinema, or medical imaging), the proposed filtering method may not be suitable. Specifically, the filtering results of the proposed method for multiview stereo images using LDIs will only be as accurate as the depth estimation and resolution of the acquired images.

6.3.4 Summary

A summary of each of the subsections mentioned above are presented in Table 6.2 for convenience.

Table 6.2: Comparative summary of the proposed 3D filter with the 2D Bilateral Filter.

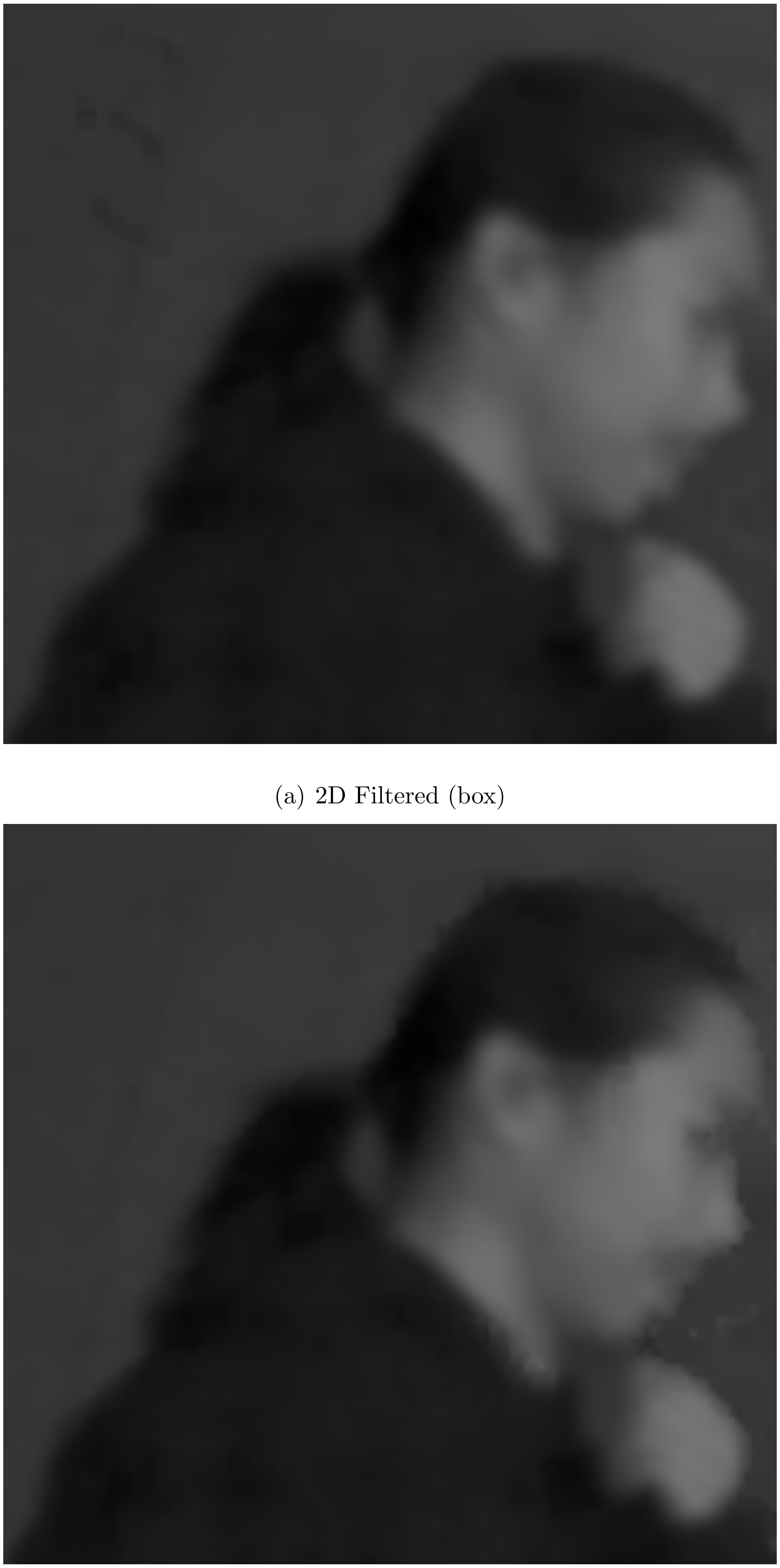
	Filtering Method	
Comparator	**2D Bilateral Filter**	**3D Filter (proposed)**
Visual: Perceived Comfort (mean out of 100)	70.6	66.5
Visual: Perceived Naturalness (mean out of 100)	64.0	60.1
Computation Time for one image with $w = 3$	48.17 sec	28.36 sec
Computation Time for four stereo pairs with $w = 3$	385.36 sec	28.36 sec
Complexity	$O(w^2 \cdot R \cdot n)$	$O(w^3 \cdot R \cdot d)$



(a) Original

(b) 3D Filtered

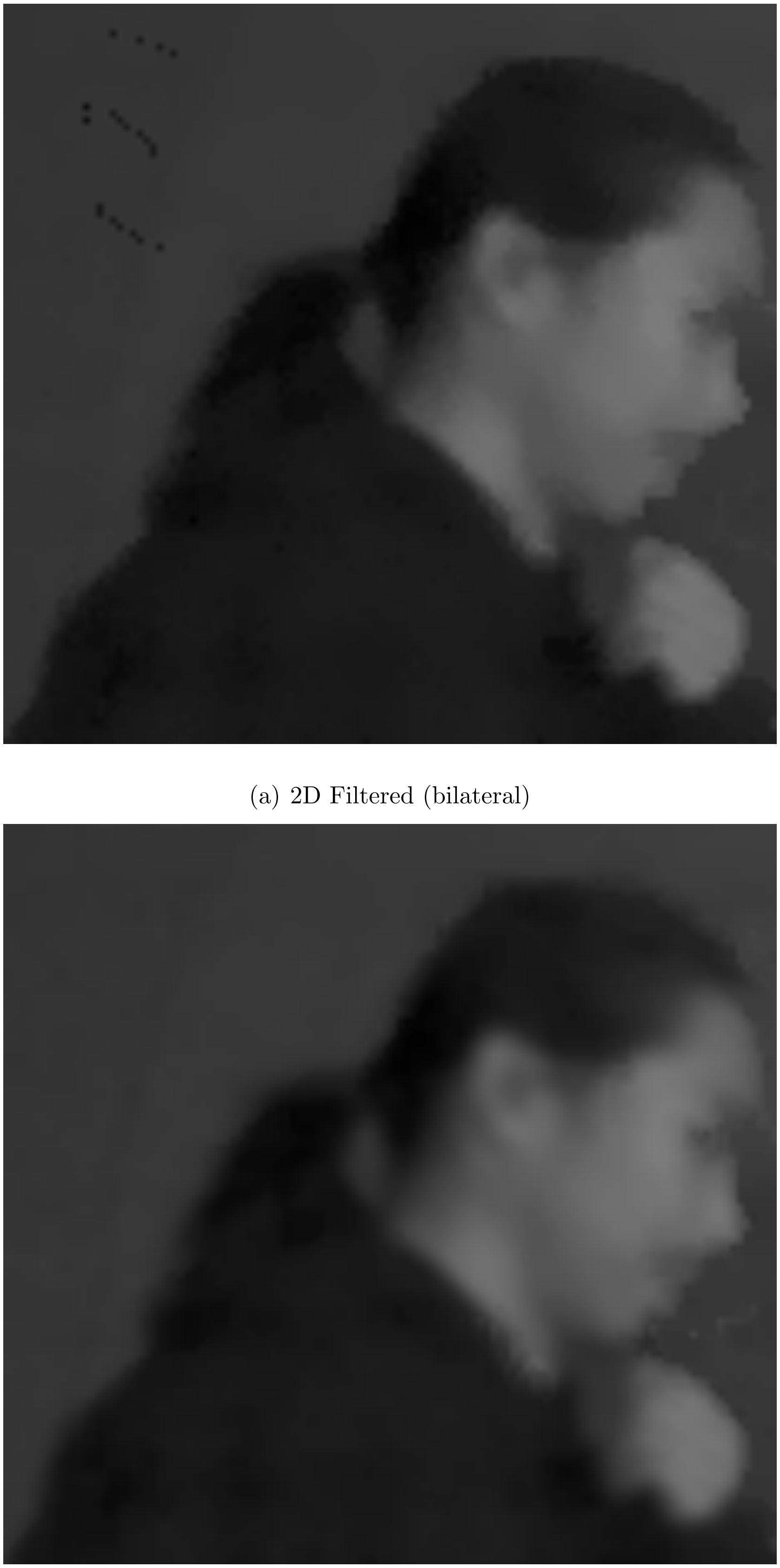
Figure 6.1: Smoothing example 1, 3D filter compared to original image with each zoomed in to show details.



(a) 2D Filtered (box)

(b) 3D Filtered

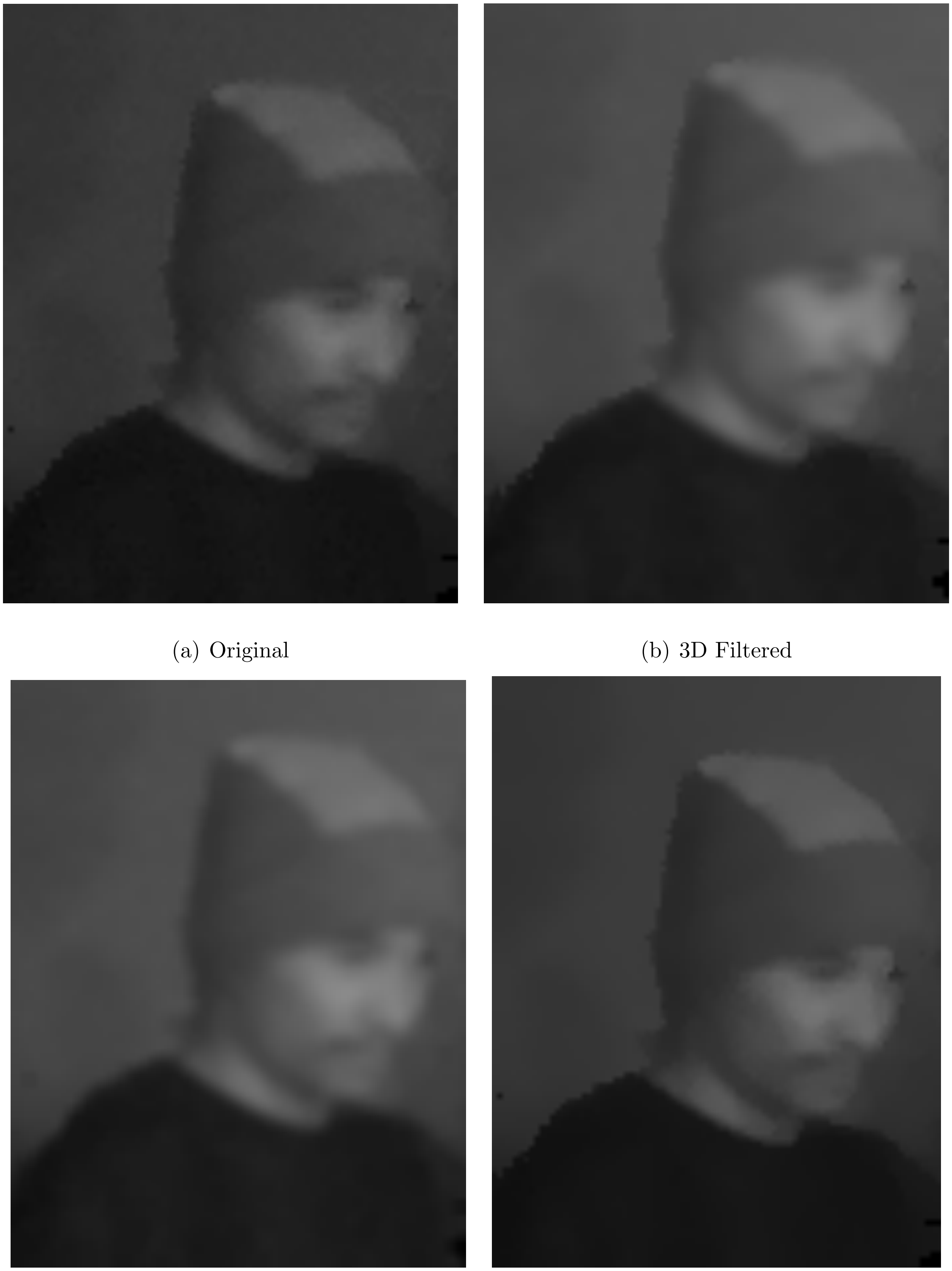
Figure 6.2: Smoothing example 1, 3D filter compared to 2D box filter with each zoomed in to show details.



(a) 2D Filtered (bilateral)

(b) 3D Filtered

Figure 6.3: Smoothing example 1, 3D filter compared to 2D bilateral filter with each zoomed in to show details.



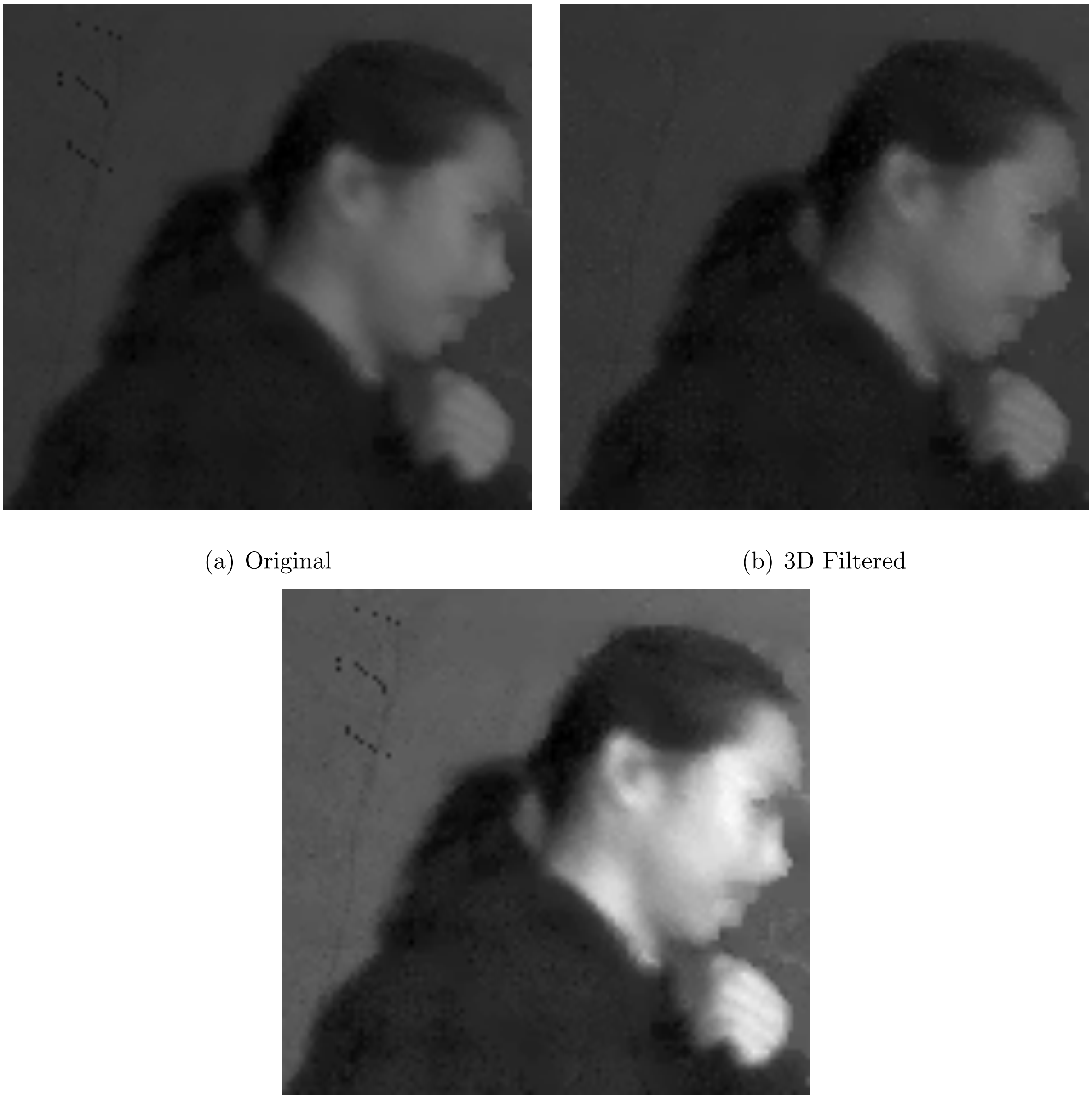
(a) Original

(b) 3D Filtered

(c) 2D Filtered (box)

(d) 2D Filtered (bilateral)

Figure 6.4: Smoothing example 2 (zoomed in to show details).

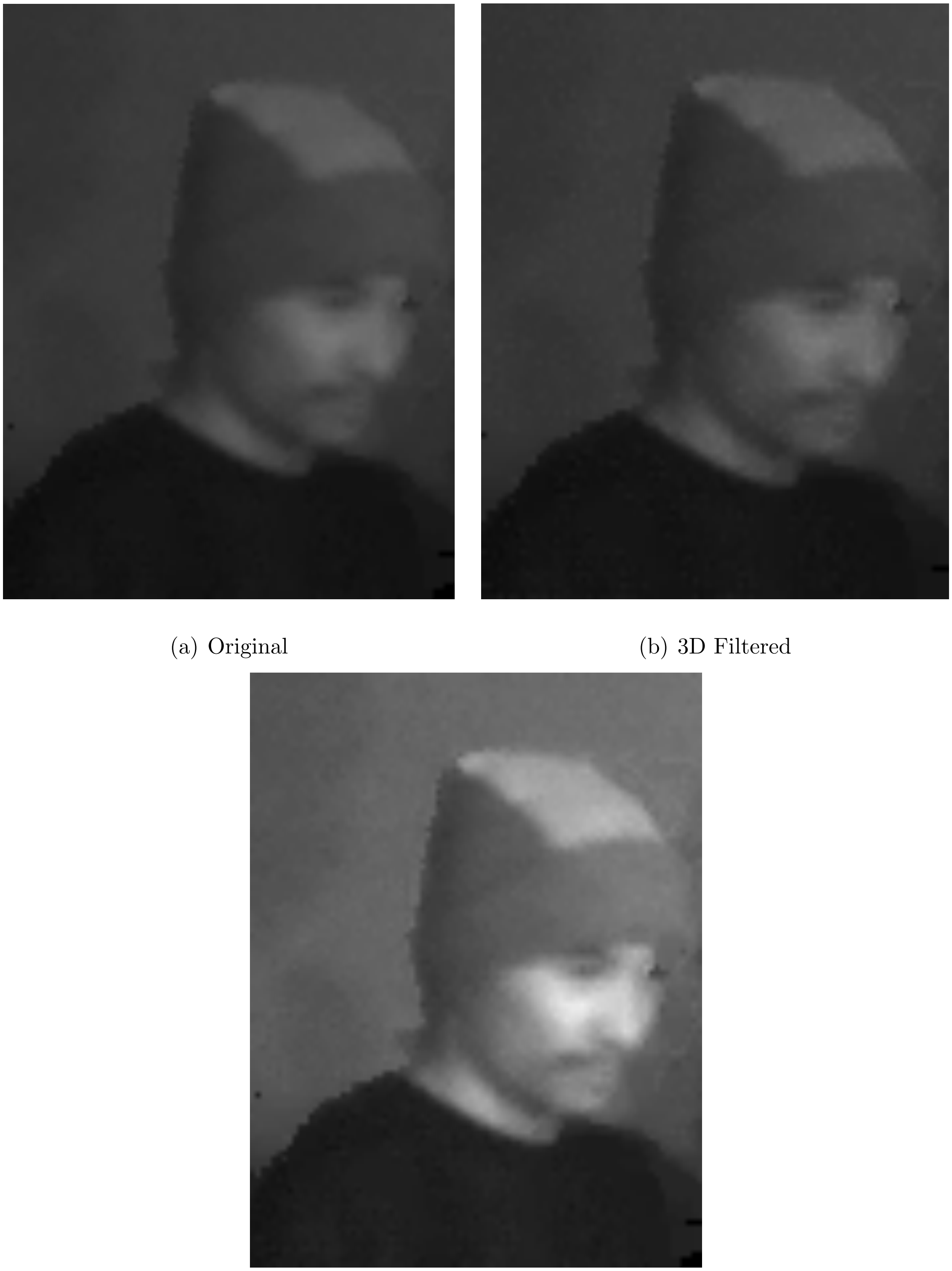


(a) Original

(b) 3D Filtered

(c) 2D Filtered

Figure 6.5: Sharpening example 1 (zoomed in to show details).



(a) Original

(b) 3D Filtered

(c) 2D Filtered

Figure 6.6: Sharpening example 2 (zoomed in to show details).

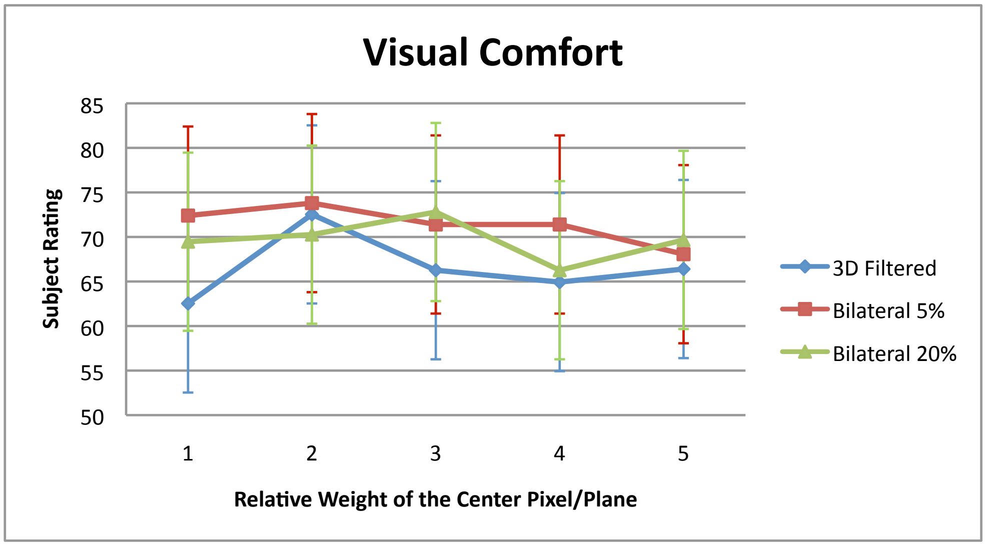


Figure 6.7: Comparison of visual comfort results of human perception.

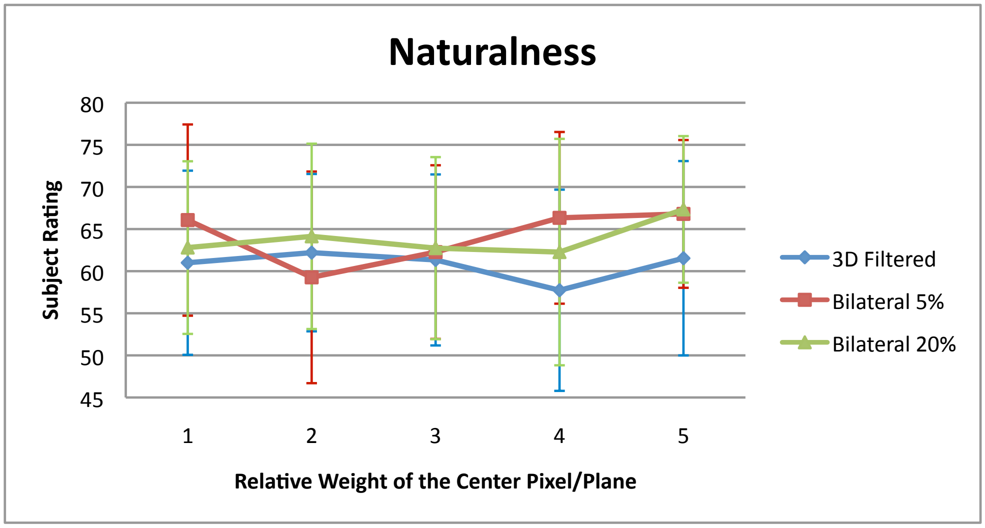
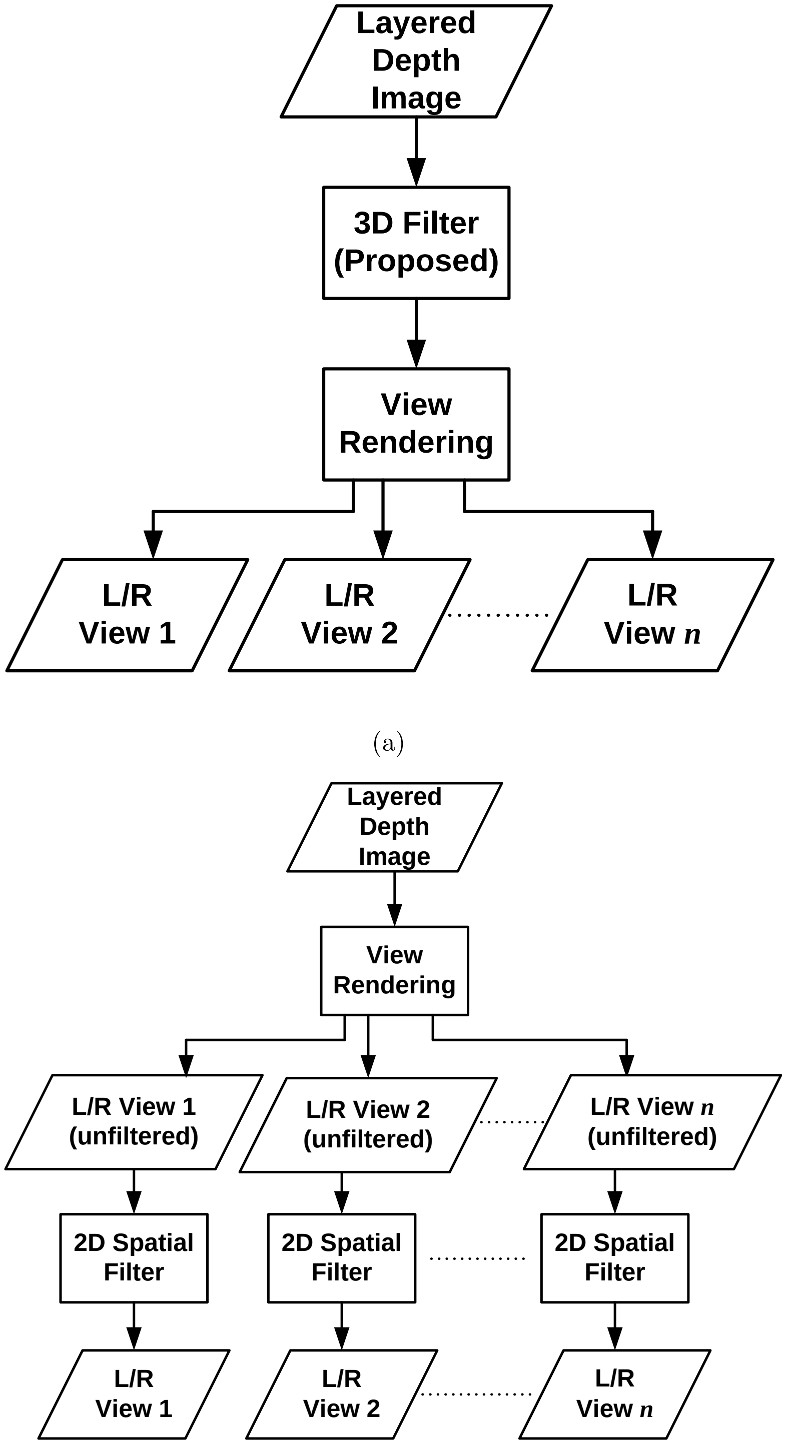


Figure 6.8: Comparison of 'naturalness' results of human perception.



(a)

(b)

Figure 6.9: Comparison of 2D and 3D LDI Filtering Algorithms. (a) 3D LDI filtering algorithm. The filter needs only to be applied once to generate any filtered view. (b) Conventional 2D LDI filtering algorithm. The filter needs to be applied n times to generate n filtered view.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This thesis presented a new approach towards filtering stereoscopic multiview images (using the LDI representation) rather than using simplistic 2D spatial filtering approaches or using the 2D bilateral filter. The method presented a 3D adaptive filter that changes size to accommodate the LDI representation. The results obtained using the proposed image smoothing and sharpening methods were similar or better when compared to the 2D filtering techniques. When comparing the image results subjectively in terms of human perception, the study conducted revealed that the proposed 3D filter can obtain similar ratings for viewing comfort and naturalness as the 2D bilateral filters, however, with much less processing time required (in a brute force comparison).

The proposed filtering method could also be used as a building block towards designing real-time filtering applications for a wide-range of multiview systems. One such system could be a synthetic depth of field for multiview images/video capable of sharpening and blurring different depths of a scene to help reduce viewing discomfort, and increase the appreciation of future multiview stereo systems. Other types of systems where this work is directly applicable are artistic applications such as virtual museums or graphics art, which will make reproduced (multiview) art appear more natural, and more comfortable.

The contribution of this thesis is a novel way to smooth or sharpen a stereoscopic multiview image using the layered depth image representation. The presented method also establishes a foundation upon which to possibly extend various 2D spatial filtering techniques into 3D for use on multiview stereo images (to be qualified by future research/investigation). This could be accomplished by modifying the filter coefficients to perform other types of filtering operations (such as edge detection).

7.2 Future Work

As mentioned earlier, stereoscopic images cause many viewers to experience discomfort. To get a better understanding of the impact this filtering method has on human perception, it would be wise to perform further experimentation with different sets of test images, and compare it against other filters (including the 3D bilateral). Additionally, it would be wise to vary the window size of the filters to determine if any point exists that causes further discomfort or leads to more natural images.

Supplementary work could also be done to apply these filtering comparisons to stereoscopic video sequences to assess their impact on quality and human perception. The experimental method presented in this work could also be used to compare other spatial operators with 3D extensions (such as other forms of sharpening, or edge detection). With a better knowledge of the human perception of these filters when applied to stereoscopic images, these filters could then be used as part of larger multiview image processing systems, and could also be used to make a synthetic depth of field system. Specifically, the filtering methods developed in this thesis could be used selectively to blur certain layers of an image, while leaving others intact, to replicate the human eye focusing on a certain depth of the scene. This would help reduce viewer discomfort in future multiview stereo systems, thus making the technology more immersive and user friendly. Other future work could be to adopt the same mechanics of the proposed filter (to adaptively change size in the LDI) to perform vector colour filtering to see if the results can be improved.

In the future, it will be useful to implement an optimized version of the proposed filtering technique to be applied in an actual real-time system using computer graphics images. The method could then be compared to the results of 3D bilateral filters and compared against optimized implementations for GPUs as in [53] and [54].

Bibliography

[1] J. Edwards, "Three-dimensional research adds new dimensions [special reports]," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 10 –13, May 2011.

[2] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07.*, June 2007, pp. 1 –8.

[3] A. Smolic, K. Mueller, P. Merkle, P. Kauff, and T. Wiegand, "An overview of available and emerging 3d video formats and depth enhanced stereo as efficient generic solution," in *Picture Coding Symposium, 2009. PCS 2009*, May 2009, pp. 1 –4.

[4] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3dtv," *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 10 –21, Nov. 2007.

[5] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Trans. Graph.*, vol. 23, pp. 600–608, August 2004.

[6] L. Onural, "Signal processing and 3dtv [in the spotlight]," *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 144 –142, Sept 2010.

[7] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3d video and free viewpoint video - technologies, applications and mpeg standards," in *IEEE International Conference on Multimedia and Expo, 2006*, July 2006, pp. 2161 –2164.

[8] H. Mitsumine, H. Noguchi, K. Enami, Y. Ninomiya, Y. Yamanoue, S. Yano, A. Hanazato, and M. Okui, "Virtual museum-3-d fine art appreciation system," *IEEE Transactions on Broadcasting*, vol. 42, no. 3, pp. 200 –207, Sep 1996.

[9] M. Okui and F. Okano, "Integral photography display for digital museum exhibit," in *Proceedings of the 9th ACM SIGGRAPH Conference on Virtual-Reality Continuum and its Applications in Industry*, ser. VRCAI '10. New York, NY, USA: ACM, 2010, pp. 365–368.

[10] J. Shade, S. Gortler, L.-W. He, and R. Szeliski, "Layered depth images," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '98. New York, NY, USA: ACM, 1998, pp. 231–242.

[11] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3dtv-a survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1606 –1621, 2007.

[12] C. Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," vol. 5291, no. 1. SPIE, 2004, pp. 93–104.

[13] A. Frick, B. Bartczack, and R. Koch, "3d-tv ldv content generation with a hybrid tof-multicamera rig," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010*, June 2010, pp. 1 –4.

[14] K. Muller, A. Smolic, K. Dix, P. Kauff, and T. Wiegand, "Reliability-based generation and view synthesis in layered depth video," in *IEEE 10th Workshop on Multimedia Signal Processing, 2008*, Oct. 2008, pp. 34 –39.

[15] B. Bartczak, P. Vandewalle, O. Grau, G. Briand, J. Fournier, P. Kerbiriou, M. Murdoch, M. Muller, R. Goris, R. Koch, and R. van der Vleuten, “Display-independent 3d-tv production and delivery using the layered depth video format,” *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 477–490, June 2011.

[16] R. Gonzalez and R. Woods, *Digital Image Processing*, 3rd ed. New Jersey: Pearson Prentice Hall, 2008.

[17] D. Morgenthaler, “Three-dimensional digital image processing,” Ph.D. dissertation, College Park, MD, USA, 1981, aAI8202854.

[18] G. Papari, P. Campisi, P. le Callet, and N. Petkov, “Artistic stereo imaging by edge preserving smoothing,” in *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009. IEEE 13th*, Jan. 2009, pp. 639 –642.

[19] Y.-R. Horng, Y.-C. Tseng, and T.-S. Chang, “Stereoscopic images generation with directional gaussian filter,” in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 30 2010-June 2 2010, pp. 2650 –2653.

[20] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heyndereckz, “Visual discomfort and visual fatigue of stereoscopic displays: a review,” *Journal of Imaging Science and Technology*, vol. 53, pp. 030 201–1–030 201–14, May-Jun 2009.

[21] W. Blohm, I. P. Beldie, K. Schenke, K. Fazel, and S. Pastoor, “Stereoscopic image representation with synthetic depth of field,” *Journal of the Society for Information Display*, vol. 5, no. 3, pp. 307–313, 1997.

[22] M. Shinya, “Post-filtering for depth of field simulation with ray distribution buffer,” in *Graphics Interface 94*, May. 1994, pp. 59 –66.

[23] G. Um, F. Speranza, L. Zhang, W. Tam, R. Renaud, L. Stelmach, and C.-H. Ahn, "Investigation on the effects of disparity-based asymmetrical filtering of stereoscopic video." SPIE, 2003.

[24] L. Stelmach, W. J. Tam, D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 188 –193, Mar 2000.

[25] J. Liu, D. Przewozny, and S. Pastoor, "Layered representation of scenes based on multiview image analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 4, pp. 518 –529, June 2000.

[26] L. Lipton, *Foundations of the Stereoscopic Cinema-A Study In Depth.* New York: Van Nostrand Reinhold Company Inc, 1982.

[27] Queen Elizabeth in 3D - on air September 20, 2010. CBC Documentaries. [Online]. Available: http://www.cbc.ca/documentaries/doczone/QE3D/

[28] MTV Live: in 3D!! - on air March 17, 2011. MTV Canada. [Online]. Available: http://www.mtv.ca/mtv-live/video_content.jhtml?id=1660140

[29] 46inch JVC XPOL Passive 3D TV Kit, RealD - The New 3D. [Online]. Available: http://www.reald.com/content/jvc-gd-463d10.aspx

[30] LG 47LD950 47inch Passive LCD 3D TV. LG. [Online]. Available: http://www.inition.co.uk/inition/dispatcher.php?URL_=product_stereovis_lg_ld920&SubCatID_=3&model=products&action=get&tab=summary

[31] G. Lawton, "3d displays without glasses: Coming to a screen near you," *Computer*, vol. 44, no. 1, pp. 17 –19, Jan. 2011.

[32] N. Baute. (2011, Mar) Surgeons use xbox to keep hands sterile before surgery. [Online]. Available: http://www.healthzone.ca/health/newsfeatures/article/960393--surgeons-use-xbox-to-keep-hands-sterile-before-surgery?bn=1

[33] S. Negry and M. First, "The effect of 2d/3d environment on decision making confidence in visual perceptual tasks," vol. 7237, no. 1. SPIE, 2009, pp. 723 716–1–723 716–13.

[34] A. Kwasinski and G. Alregib, "Gadgets and signal processing: The 2010 international ces [in the spotlight]," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 176 –170, May 2010.

[35] J. Edwards, "Signal processing plays prominent role at 2011 international ces [special reports]," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 13 –16, May 2011.

[36] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3dtv services providing interoperability and scalability," *Signal Processing: Image Communication, Special issue on three-dimensional video and television*, vol. 22, no. 2, pp. 217 – 234, 2007.

[37] P. Fua, "A parallel stereo algorithm that produces dense depth maps and preserves image features," *Machine Vision and Applications*, vol. 6, pp. 35–49, 1993.

[38] S. Battiato, S. Curti, M. L. Cascia, M. Tortora, and E. Scordato, "Depth map generation by image classification," vol. 5302, no. 1. SPIE, 2004, pp. 95–104.

[39] Y.-S. Huang, F.-H. Cheng, and Y.-H. Liang, "Creating depth map from 2d scene classification," in *3rd International Conference on Innovative Computing Information and Control, 2008. ICICIC '08.*, June 2008, p. 69.

[40] Y. Matsumoto, H. Terasaki, K. Sugimoto, and T. Arakawa, "Conversion system of monocular image sequence to stereo using motion parallax," vol. 3012, no. 1. SPIE, 1997, pp. 108–115.

[41] Y.-L. Chang, C.-Y. Fang, L.-F. Ding, S.-Y. Chen, and L.-G. Chen, "Depth map generation for 2d-to-3d conversion by short-term motion assisted color segmentation," in *IEEE International Conference on Multimedia and Expo, 2007*, July 2007, pp. 1958 –1961.

[42] P. V. Harman, J. Flack, S. Fox, and M. Dowley, "Rapid 2d-to-3d conversion," vol. 4660, no. 1. SPIE, 2002, pp. 78–86.

[43] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. 2003*, vol. 1, June 2003, pp. I–195 – I–202.

[44] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," *International Journal of Computer Vision*, vol. 35, pp. 151–173, 1999.

[45] L. Do, S. Zinger, and P. H.N. de With, "Quality improving techniques for free-viewpoint dibr," vol. 7524, no. 1. SPIE, 2010, pp. 75 240I–1 – 75 240I–10.

[46] K.-J. Oh, S. Yea, and Y.-S. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-d video," in *Picture Coding Symposium, 2009. PCS 2009*, May 2009, pp. 1 –4.

[47] C.-M. Cheng, S.-J. Lin, S.-H. Lai, and J.-C. Yang, "Improved novel view synthesis from depth image with large baseline," in *19th International Conference on Pattern Recognition, 2008. ICPR 2008.*, 2008, pp. 1 –4.

[48] X. Cheng, L. Sun, and S. Yang, "Generation of layered depth images from multiview video," in *IEEE International Conference on Image Processing, 2007. ICIP 2007.*, vol. 5, 16 2007-Oct. 19 2007, pp. 225–228.

[49] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplar-based inpainting," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2, 2003, pp. 721–728.

[50] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "A gentle introduction to bilateral filtering and its applications," in *ACM SIGGRAPH 2007 courses*, ser. SIGGRAPH '07. New York, NY, USA: ACM, 2007.

[51] A. S. Babalis, A. N. Venetsanopoulos, and D. Androutsos, "Investigation of the effect of three-dimensional smoothing on multiview stereo images," in *17th International Conference on Digital Signal Processing, Corfu, Greece, July 6-8, 2011 (accepted)*, July 2011.

[52] A. S. Babalis, A. N. Venetsanopoulos, and D. Androutsos, "Smoothing of stereoscopic images for artistic imaging applications," in *Proceedings of the Electronic Imaging and the Visual Arts (EVA2011) Conference, Training and Workshop, Florence, Italy, May 4-6, 2011*, May 2011, pp. 48–53.

[53] J. Chen, S. Paris, and F. Durand, "Real-time edge-aware image processing with the bilateral grid," in *ACM SIGGRAPH 2007 papers*, ser. SIGGRAPH '07. New York, NY, USA: ACM, 2007.

[54] A. Adams, N. Gelfand, J. Dolson, and M. Levoy, "Gaussian kd-trees for fast high-dimensional filtering," in *ACM SIGGRAPH 2009 papers*, ser. SIGGRAPH '09. New York, NY, USA: ACM, 2009, pp. 21:1–21:12.